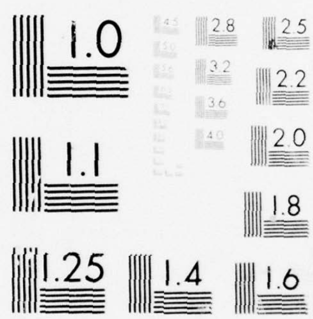
AD-A040 885 RAND CORP SANTA MONICA CALIF F/G 5/1
TOWARD MORE EFFECTIVE USE OF EXPERT OPINION: PRELIMINARY INVEST--ETC(U)
OCT 76 H SACKMAN
UNCLASSIFIED P-5570 NL

1 OF 1
AD
A040885

END
DATE
FILMED
7-77



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

6

TOWARD MORE EFFECTIVE USE OF EXPERT OPINION:
PRELIMINARY INVESTIGATION OF PARTICIPATORY
POLLING FOR LONG-RANGE PLANNING

10 H. Sackman

11 October 1976

12 86p.

14 P-5570

296600

PREFACE

The research reported in this paper was performed in the course of an evaluation of methods of long-range planning for Air Force Research and Development. I am most indebted to Giles Smith, who contributed substantively to the design of the questionnaire used in this study, to the selection and briefing of expert panelists, and to the interpretation of results. The continuing interest and helpful efforts of Ed Ojdana in developing and explicating this technique are most appreciated. Particular gratitude is due other Rand Colleagues who participated in the experimental demonstration participatory polling.

This publication reports the results of a research experiment intended merely to examine the feasibility and methodology of participatory polling of experts. The results of the demonstration questionnaire relating to military findings are entirely illustrative; in no sense do they represent Rand recommendations or opinion, and they do not bear on or derive from Air Force policy. Instead, they reflect the opinions, at the time, of a small number of Rand staff members who agreed to participate in a polling experiment and who had some background knowledge of the topic being considered. Similarly, the research recommendations set forth at the end of this paper reflect only the personal views of the author, and do not necessarily represent the views of others at Rand.

CONTENTS

TABLES

FIGURES

1. BACKGROUND

Everyone seeks and uses expert opinion in virtually all walks of life. This applies to organizations and institutions as well as to individuals. Personal and group decisions are often based on perceptions of expert consensus. With growing social complexity and an accelerating tempo of contemporary social change, society is committing itself more deeply to expert opinion for planning, problem-solving, guidance and reassurance.

1.1. The Scientific Challenge of Expert Opinion

The old authoritarian roles of the past have been supplanted by new technicians and specialists, the contemporary experts for social affairs, who increasingly are called upon to dispense their expertise by opinion and intuitive fiat, without the benefit of scientific checks and balances. This lag between authoritarian social practice based on opinion, and scientific method based on empirical experimentation, has contributed to a kind of social schizophrenia where science contributes freely to technological affairs, and is indifferent to the working basis of human decisions increasingly determined by arbitrary use of expert opinion.

Social scientists have made major strides in methodology and findings concerned with human attitudes, the psychology of judgment, and opinion polling. The history of this development has been coextensive with empirical test and analysis of individual differences going back to Galton's pioneering studies a century ago. A rigorous empirical methodology has evolved for psychometric techniques and opinion questionnaires including testing theory and questionnaire design, item analysis, sampling reliability, quantitative scaling and scoring procedures, and replicable validation procedures. This methodology has been applied to countless applied tests, and a code of approved ethical and scientific testing procedures and practices has been put forth by professionals seeking highest contemporary standards to minimize social abuse and encourage social acceptance, (e.g., American Psychological

Association, 1974). Similar standards have evolved for social experimentation with group process to facilitate objective evaluation of group decision-making and group problem-solving.

This introduction is aimed at three key issues:

1. Increasing use of expert opinion of all types is endemic at all levels of our society, fundamentally without the benefit of scientific method.
2. For various historical reasons, there is virtually no rigorous scientific work on the nature and extent of social reliance on expert opinion even though many relevant standards, procedures, and findings have been developed in related areas.
3. The scientific study of the use of expert opinion is long overdue, not only for the basic knowledge it can provide on the dynamics of social decision-making and problem-solving, but for the great promise of scientific method in discriminating between misuse of expert opinion, and more socially beneficial applications of expert opinion.

1.2. Delphi and Expert Opinion

The above issues were developed from a recent and extensive critique of the Delphi method conducted by the author at Rand (1975). Although there are many variations of Delphi that have appeared in the literature, conventional Delphi is basically an interative opinion questionnaire technique, with anonymous feedback after each iteration, applied to a panel of experts until "optimal" consensus is reached among the panelists. Rand has been involved in developing policy sciences since it first appeared as a nonprofit corporation in 1948. Delphi was developed at Rand, largely by Helmer and Dalkey. Since a well-publicized study in 1963 (Gordon and Helmer), it has spread rapidly in government and industry, not only in the United States, but also throughout the world. Over 1,000 Delphi studies have appeared in a

highly diversified applications literature. Nevertheless, the Delphi technique was a continuing source of controversy at Rand where it was felt that a comprehensive evaluation should be performed. As expected, the evaluation of Delphi also turned out to be controversial. The main findings of this evaluation, relevant to this study, are outlined below.

Delphi responded to a deeply felt need, particularly in corporate and government communities, for an apparently rational, systematic, easily understood, fast, convenient, and inexpensive method to solicit expert opinion and determine expert consensus in virtually any area of interest. The remarkable growth of this technique in widely disparate fields for many different types of users attests to growing social needs for expert opinion. For example, in a survey of 65 corporations, Hayden (1970) found that 69 percent of these companies used expert panel techniques for assistance in technological forecasting, with 26 percent reporting the use of Delphi.

The review of the literature associated with Delphi indicated a lack of interest in scientific evaluation of expert opinion on the part of social scientists. For example, in economics, Zarnowitz (1965) found that simple arithmetic extrapolation of the gross national product from 1953 to 1963 was not significantly different from the average accuracy of expert opinion in eight independent forecasts. Slovic (1972) found similar results for various types of market statistics. In psychiatric and psychological diagnosis and prognosis, Hersch (1969), and Goldberg (1968) reported basically negative and contradictory findings in professional clinical opinion studies. In applied psychology, Thorndike (1949) reported extensive World War II experience demonstrating the superiority of standardized paper-and-pencil tests over professional opinions of interviewers in large military training programs. There were no systematic, continuing studies of expert opinion in Berelson and Steiner's (1964) inventory of findings on human behavior, or in the International Encyclopedia of Social Science (Sills, 1968). These citations support the earlier assertion of a lack of interest in social science for studies of expert opinion, and helps explain the

great popularity of Delphi as a pseudo-scientific technique in filling this void.

On the methodological side, Heller (1969) has reviewed the advantages and limitations of variations of the mailed questionnaire versus face-to-face interviews. He has concluded that rigorous eclectic methods are probably the preferred way to get the best of both worlds (e.g., "hard statistical data and soft insights"). His proposal for "group feedback analysis," based on a type of iterative polling, is, like Delphi, a precursor of the participatory polling technique described in this paper.

The author's evaluation of Delphi led to a recommendation that, except as an exercise for informal heuristic purposes, conventional Delphi should be dropped from institutional, corporate and government use until its principles, methods, and fundamental applications can be established as scientifically tenable. The difficulties in the scientific use of Delphi are common to most procedures involving expert opinion. These are very briefly mentioned to illustrate key methodological problems in this area.

There are a whole series of issues related to the design and development of questionnaires applied to experts. The questionnaire should be based on a thorough review of the relevant literature and on some form of construct validity for the particular area being investigated, and should represent systematic sampling of items within this area. The questionnaire should undergo pilot testing and item analysis, utilizing standard statistical procedures to weed out unreliable, invalid, and redundant items. Such safeguards for item reliability and validity were rarely followed in Delphi studies.

Another series of issues was related to the selection of expert samples. As in any replicable testing effort, sample parameters need to be explicitly stated in quantitative form, including basic information such as age, sex, socio-economic status, educational level, professional training and experience, and specification of the collection of skills meeting operational definitions that distinguish experts from nonexperts. These requirements were virtually ignored in expert

sampling procedures in Delphi studies, and in many cases were deliberately omitted, in accordance with the Delphi requirement for anonymity of respondents. If the expertise of panels is not specified and measured, replication is impossible.

Then there are a variety of technical issues associated with the use of experts in Delphi and in expert opinion studies broadly considered. A crucial question is whether the "experts" respond to questions in a manner distinguishable from nonexperts. There is a significant experimental literature indicating that, for complex and wide-ranging social issues, predictions of teachers are the same as students (McGregor, 1938, and Cantril, 1938), short-term forecasts of more knowledgeable individuals are not any better than those less knowledgeable (Kaplan et al, 1950), short-term economic forecasts of more experienced professionals are indistinguishable from those less experienced (Campbell, 1966). Reisman et al (1969) found similar Delphi results between experts and laymen in estimating the value of social services, and Bedford (1972) found that housewives gave virtually the same responses as professionals in forecasting telecommunication services for the home. These and related results suggest that the burden of proof be placed on the investigator to demonstrate that his sample of "experts" can provide opinions significantly and usefully better than the opinions of nonexperts.

Other technical issues include safeguards against expert halo effects. "Experts" are subject to group suggestion, including overt and covert manipulation of opinion by investigators, as demonstrated many years ago by Sherif (1936) in his experiments on the autokinetic effect. The iterative Delphi procedure, aimed at group consensus, is quite similar to social psychological techniques developed in studying group conformity, rejection of deviant opinion, and deindividualization, all of which have been shown to be counterproductive in influencing the quality of group decisions.

Anonymity in expert group process can contribute to personal unaccountability on the part of the panel, the investigator, and the user of results. Anonymity and associated unaccountability in expert sampling procedures can lead to elitism and deliberate manipulation of

results to satisfy vested interests. A recent Delphi study reported by Brodeur (1973) in the New Yorker illustrates this point. Almost all of the panel of medical experts--to assess the maximum allowable level of asbestos particles to avoid asbestos-induced cancer--were selected from the ranks of paid medical consultants for the asbestos industry. Scientists are needed to develop techniques and improve standards for the use of expert opinion to minimize such abuses of the public interest.

An additional cluster of issues is concerned with expert behavioral inputs leading to expert outputs. Available evidence from Delphi studies reveals snappy "expert" answers to ambiguous questions. Response time to individual items on lengthy questionnaires may frequently be less than 10 seconds, including time for reading and recording answers. This type of virtually instantaneous information processing is hardly the basis for reasoned and balanced expert opinion. Yet, unpaid experts in a hurry, reinforced by reassurances from investigators that *responses to questionnaires* require very little personal time, can grind out responses at a very rapid rate.

Expert behavior needs to be directly studied and analyzed in its own right, including personal information processing procedures, time and motion studies, types of data bases used, and on-the-spot behavioral protocols, if we are to develop confidence in at least some forms of expert opinion. Expert thinking need not be regarded as an intuitive black box, inaccessible to scientific analysis.

Compounding these difficulties are amorphous questionnaire items with minimum qualification of terms, baseline data, or background conditions and assumptions. As such, they often represent inkblots that permit experts to project many different scenarios and types of interpretations which make for noncomparable responses with runaway increases in error variance.

This summary of the difficulties in obtaining scientifically reliable and valid Delphi results applies directly to research in expert opinion broadly considered, and indicates some of the directions necessary for much more rigorous research. The Delphi experience has

been extremely useful in pointing up pitfalls and limitations in harnessing expert opinion and moving on to more demonstrably potent procedures for more reliable and valid results.

2. BASIC PRINCIPLES OF PARTICIPATORY POLLING

The need for this study arose from two sources. First, a comprehensive review of Air Force and industrial practices in long-range R&D planning indicated that the most desirable and useful types of planning require substantial participation of technical experts, planners, and decisionmakers in soliciting their assumptions, values, and opinions. Second, current techniques for obtaining such inputs, such as Delphi, as described earlier, suffer from numerous and severe limitations that need to be overcome. In response to these needs, the technique of participatory polling was developed for a prototype trial which constitutes the basis of this study.

The design of a new technique for experts was based on the eight objectives listed below.

1. Expert-Interpreted Opinion. Obtain systematically interpreted opinion from experts, policy makers, and other selected populations in specified areas of inquiry.
2. Balanced Adversary Procedure. Determine level of observed consensus and identify polarization of attitudes, in a balanced adversary approach that provides equal treatment for all participants whether they elicit conformist or outlier responses.
3. Shared Knowledge of Results. Provide shared opinions and supporting rationale through iterative exposure of results.
4. Group Data Base. Develop a data base of pooled group opinion.
5. Participatory Evaluation. Solicit active participation of subjects to obtain evaluations and explanations of expressed opinion throughout the entire study cycle.
6. Flexible Anonymity. Participants generally remain individually anonymous in iterative responses to the structured questionnaire to minimize undue influence associated with personal halo effects. However, depending upon explicit agreements set out prior to the initiation of the inquiry,

direct interaction between part or all of the group may be designed into the technique with full knowledge of the entire group.

7. Scientific Standards. Meet established scientific standards for questionnaire development and use in an iterative experimental setting with selected participants. Heuristic exercises should be clearly distinguished from formal, experimental studies.
8. Cost-Effective Technique. Develop a versatile, cost-effective technique that builds upon the lessons of the Delphi experience and related areas for a substantially improved method for systematically eliciting participant-interpreted opinion.

Given the above eight objectives, the new technique was described as participatory polling, in recognition of the contribution of expert participants in generating, explicating, and evaluating their own opinion, and the opinions of others. The immediate objective of this study was to apply this new technique, for the first time, to a substantive Air Force policy area in long-range planning.

The selected area was Air Force R&D planning for the close air support mission. It was felt that this planning task was neither too broad (as with, for example, strategic Air Force planning) nor too highly specialized (as with advanced electronic jamming countermeasures). The statement of the study problem is essentially the initial exploratory application of participatory polling to long-range Air Force R&D planning objectives for the close air support mission. The next section describes the operational procedures developed to implement this prototype demonstration of participatory polling.

3. PROCEDURE

Methodology covers three basic areas: objectives, questionnaire design, and experimental design. Objectives refer to the scope and procedural goals of the study; experimental design refers to empirical procedures, performance measures and formal hypotheses tested by statistical techniques; questionnaire design covers definition of the domain of inquiry and its translation into formal questionnaire format. Each of these three areas is discussed below.

3.1. Methodological Objectives

This study was conceived as an exploratory investigation of participatory polling conducted in the spirit of a prototype demonstration. This informal effort is not put forth as a formal experimental study in its own right. The main methodological objectives are to explore the feasibility, acceptability, credibility, and gross cost-effectiveness of an initial application of participatory polling.

Secondary methodological objectives involve exploration of the nature and dynamics of expert opinion in this prototype study setting. These include such characteristics as individual differences between experts, areas of relative agreement and disagreement, reasons for opinions offered, and the types of adversary positions developed in response to the iterative polling procedure.

A third group of methodological objectives is concerned with psychometric characteristics of participatory polling in general, and for this questionnaire in particular. These include considerations of test reliability, various aspects of test validity, approaches to item analysis, and special hypotheses unique to the procedure such as opinion changes associated with iterative application, and the reliability of ratings for reasons supporting expressed opinions.

The final group of methodological objectives is concerned with illustrative findings for long-range Air Force planning. Such objectives involve the perceived value of the participatory polling exercise in facilitating planning, ordering priorities, generating insights, and organizing the domain of inquiry in a rational, consistent manner for the policy-maker. The goal of these objectives is to determine

whether follow-on developments are desirable, and if so, the form and directions such developments should take. Thus, the basic goals of this study are illustrative and methodological; in no sense was this study designed to arrive at substantive policy recommendations for long-range Air Force planning for the close air support mission.

3.2. Questionnaire Design

Within the framework of the eight objectives of participatory polling cited earlier, the basic task in questionnaire design was to develop a logical set of items that would encompass operational requirements of the close air support mission in a systematic manner. These questionnaire items would have to be suitable for submission to an expert panel to solicit their opinions and supporting reasons in a balanced adversary setting. This task is fundamentally a problem in content and construct validity for questionnaire design. Content validity is concerned with defining the domain of inquiry, and with representative sampling of items to cover the domain adequately in the final questionnaire. To take a simple example, an achievement questionnaire on world geography should not be exclusively concerned with countries only in the Northern Hemisphere. Construct validity is concerned with the theories, hypotheses, and concepts concerning the substantive domain (the close air support mission for R&D planning).

Basically, the scheme developed by Smith was to logically partition the range of possible scenarios for the close air support mission in accordance with five categories. These include three environmental and two operational categories. The environmental variables are type of war (NATO versus Third Area); visibility conditions (clear day versus restricted visibility); and type of penetration (manned penetration versus nonpenetration, e.g., piloted aircraft versus remotely piloted vehicle). The two operational variables include finding the target (fixed versus mobile targets); and coping with defenses (guided weapons versus guns).

Note that each of the five close air support categories are dichotomized, making a maximum possible total of 2^5 or 32 logical

combinations where each "scenario" constitutes a single questionnaire item. The items concerned with survival of unmanned vehicles were not considered as critical as the other items, and were consequently dropped, leaving a total of 24 items for the questionnaire. For example, the first item of the questionnaire was a close air support mission which required "coping with missiles, in a NATO war, on a clear day, with manned penetration." The complete listing of the 24 questionnaire items is shown in Exhibits A and B, including definition of terms, supporting rationale, and specification of the range of scenarios included.

The object of the first round was to obtain importance ratings for each of the 24 items from each expert in the panel, and the "main reason" for the importance rating given, typically in the form of a single phrase or sentence. The four-step importance scale ranged from: 0, unimportant; to 1, marginally important; to 2, important; to 3, very important. Complete instructions for the first round are shown in Exhibit A.

In the second round, participants were given the same questionnaire with feedback, approximately one month after the first round. The feedback included the following: descriptive statistics on group response to the first round (mean and standard deviation of importance ratings for each item including the frequency distribution of ratings); and a list of three to five reasons for each item, including "pro" and "con" adversary positions derived from first-round open-end responses. The experts were asked to provide new ratings for each item based on all available information from the expert panel, and to rate the importance of each of the listed reasons shown for each item. This procedure ensured examination and consideration of the reasons and ratings offered by others before second-round ratings were entered for the final results. Thus, the second round provided final item ratings and evaluations of first-round reasons. Exhibit B includes the complete second-round questionnaire, showing detailed instructions, first-round feedback, and adversary reasons for ratings.

3.3. Experimental Procedure

The experimental procedure is discussed under three categories, subjects, questionnaire administration, and statistical design. The subjects refer to the Rand expert panel, test administration includes preparation, briefing, and written instructions for the participants, and statistical design includes basic experimental design and exploratory statistical testing.

3.3.1. Subjects. The subjects were 10 Rand professionals, all with extensive careers in the analysis of military problems. They were personally solicited on a voluntary basis by their colleague, Giles Smith, who is principal investigator of these Project Rand studies on long-range Air Force planning for research and development. No formal sampling measures were obtained as to education, years of experience, type of experience, or other population parameters in view of the informal, exploratory nature of this preliminary study. Formal experimental studies should have full data on expert sampling characteristics relevant to the study setting.

The subjects were solicited by way of an initial telephone request and a group meeting. At the group session, lasting approximately an hour and a half, the background of this Project Rand study was reviewed and instructions for the questionnaire were discussed. All panelists had ample opportunity to raise any questions they wished, or to decline participation in the study. Only a few declined. For the basic objectives of this initial demonstration, the obtained sample of 10 experts is sufficient.

3.3.2. Test Administration. During the initial briefing, considerable time was spent on defining the close air support mission as specified by the questionnaire items, and fielding any questions raised by the panel. (See Exhibit A for specification of the item scenarios, and the rating scale.) We felt that this discussion was essential for understanding the ground rules of the substantive military content of the questionnaire, and for stimulating motivation among the respondents by encouraging them to participate personally in the initial open discussion. Motivation was also enhanced by indicating that full results in the final report would be made available to all participants.

The crucial importance of providing written reasons for each mission rating was heavily stressed as the unique characteristic of this study. Verbal justification of each first-round rating was mandatory, not optional. These open-end responses provided the basis for the adversary reasons supporting the spread of observed ratings for each questionnaire item. (These reasons are listed in Exhibit B for the second-round questionnaire.)

Although the group meeting made total anonymity impossible since each member of the panel knew who the other members were, we nevertheless retained basic anonymity of individual responses in pooled group feedback. Thus, although each individual was aware of his own quantitative and qualitative responses, he did not know how any other particular individual responded unless he was sufficiently motivated to inquire on his own. We had no objections to any participant putting in extra time and effort to compare his own views with those of others if he so wished, primarily because we believed that more discussion would generally lead to more useful and thoughtful opinion.

At the briefing, and in the written questionnaire instructions, it was emphasized that this exercise was aimed at a balanced adversary process for the full range of expert opinion as opposed to deliberate generation of consensus. The justification for this approach was our belief that the policy maker should have the benefit of the full spectrum of adversary positions and supporting rationale in arriving at his own decisions. Accordingly, in contrast to alternative techniques, such as conventional Delphi, the experimental procedure treated all panelists equally, whether or not they were "mavericks" or "conformists."

The second round of the questionnaire was administered approximately one month after the first round. This was about the minimum interval possible because of "stragglers," and the requirement to analyze statistical and verbal responses for second-round feedback after all participants returned their completed questionnaire. The analysis was facilitated by keypunching all numerical and verbal data, storing the data on magnetic tape, and selectively printing out

specified subsets of data as required for the analysis. Even with the small sample of only 10 subjects, computerization greatly facilitated all analyses compared to cumbersome manual sorting and shuffling through the original questionnaire forms.

The second-round questionnaire was handled entirely through inter-office mail at Rand and did not require any special briefing. The instructions requested that each participant rate the various reasons offered by the panel for each item and then provide his own final rating for the item. In addition, an "Overall Questionnaire Evaluation" form was filled out, providing various types of information on reliability, validity, and problems with the questionnaire. Opportunity was provided for additional comments, if desired, for individual items and for the overall questionnaire. Individual anonymity was assured for the second as well as the first round. A copy of the panelist's complete first round questionnaire was made available for reference purposes in working on second-round responses. See Exhibit B for the complete second-round questionnaire.

3.4. Statistical Design

There are two types of statistical analyses that were planned for this exploratory study. The first covers various types of descriptive statistics pertinent to this two-round exercise. The second is concerned with special statistical tests constituting the experimental design of the study and item analyses of the questionnaire. Each is discussed in turn.

3.4.1. Descriptive Statistics. The basic descriptive statistics refer to results for items, reasons, and overall evaluation. The descriptive statistics include means and standard deviations for importance ratings for each of the 24 questionnaire items for the first and second rounds. For panelist reasons, the descriptive statistics include means and standard deviations of ratings for each separate reason listed for each of the 24 items for the second round. These varied from three to five reasons per item. For "overall evaluation," the descriptive statistics are raw score frequency distributions for the

three objective items in the evaluation portion of the second-round questionnaire. All of these three types of descriptive statistics are found in Exhibit B.

3.4.2. Special Statistics. Special statistical tests were applied to the basic experimental design of this study and to various aspects of test reliability. The experimental design is concerned with three key variables in participatory polling: 1) subjects (S), 2) rounds (R), and 3) items (I). In this study there are 10 subjects, 2 rounds, and 24 questionnaire items. Thus, this experimental layout can be tested by a three-way analysis of variance, S versus R versus I, with one observation per cell. Subjects are not pooled together because the analysis of differences between experts is of major importance in its own right. This statistical design provides tests for the three main effects (S, R, and I), and three interaction effects (SR, SI, and RI). The three-way interaction effect (SRI) is used as the error term in determining statistical significance. Thus, the basic experimental design permits tests for differences in ratings between subjects, rounds and items, and for second-order interactions between these three variables.

Two types of questionnaire reliability are examined by additional statistical tests. The first is a straightforward test-retest reliability between the first and the second round. Specifically, the ratings of each expert for each of the 24 items on the first round is correlated against the same ratings for second round conditions approximately one month later. This procedure provides one type of test-retest reliability for each of the 24 items.

The second reliability test is concerned with reasons put forth in support of importance ratings for each item. The hypothesis is that experts will generally be able to discriminate some of the reasons as being more important than other reasons for a given mission scenario item. If this is the case, then the mean importance ratings for the various reasons should be significantly different from each other according to an analysis of variance test separately conducted for each item.

In closing this section, it should be noted that the various descriptive statistics and special statistical tests comprise a framework for formal item analysis of the questionnaire. This item analysis includes tests of concurrent validity through measures of internal consistency, and tests of item reliability.

4. PILOT STUDY RESULTS

The results essentially cover first-round versus second-round outcomes, analysis of variance findings, and open-end evaluations, treated in the same order. As mentioned earlier, Exhibit B contains first-round and second-round descriptive statistics, including means and standard deviations for all 24 items, and for all listed reasons in support of item ratings. The reader should refer to Exhibit B which lists these results in the questionnaire context.

4.1. Comparison of First-Round Versus Second-Round Ratings

Table 1 compares importance ratings for the first and second rounds in rank order of mean item ratings. That is, the highest rated item is at the top and the lowest (for the first round) is at the bottom. The rank order for mean item ratings is separately shown for the two rounds. (Tied means were randomly assigned successive ranks.) The standard deviation associated with each mean item rating is shown in parentheses in Table 1. The last column in this table is the product-moment correlation of ratings for round one versus round two for the 10 subjects for each item.

Thus, the first row in Table 1 is read as follows: "Finding mobile targets, in a third area war, on a clear day, for a manned penetration mission," was the item-scenario that received the highest importance rating for close air support R&D for both rounds; the mean of 2.6 for round one and 2.7 for round two is closest to the top category of the rating scale, "very important," which is a rating of 3. Note that the standard deviation substantially decreased from round one to round two for this item (from .70 to .48). Note also that the test-retest reliability for this item is a Pearson product-moment coefficient of .92.

Several characteristics of Table 1 are noteworthy. A quick scan of rank order of mean ratings from round one to round two indicates rather close correspondence. The Spearman rank-order correlation between round one and round two is .92 for the 24 items with a standard error of .04, indicating a very high and reliable stability of ratings from round one to round two.

Table 1

MEAN IMPORTANCE RATINGS FOR CLOSE AIR SUPPORT ITEMS FROM ROUND ONE TO ROUND TWO

Item	Rank Order First Round	Rank Order Second Round	Mean Ratings First Round	Mean Ratings Second Round	Correlation R1 vs. R2
Finding Mobile Targets: Third Area, Clear Day, Manned Penetration	1	1	2.60 (.70)	2.70 (.48)	.92
Finding Mobile Targets: NATO War, Restricted Visibility, Manned Penetration	2	3	2.50 (.71)	2.60 (.52)	.91
Coping With Missiles: NATO War, Clear Day, Manned Penetration	3	2	2.40 (.70)	2.60 (.52)	.80
Finding Mobile Targets: NATO War, Clear Day, Manned Penetration	4	5	2.30 (.82)	2.50 (.53)	.90
Coping With Missiles: NATO War, Restricted Visibility, Manned Penetration	5	4	2.30 (.82)	2.60 (.52)	.84
Finding Mobile Targets: Third Area, Restricted Visibility, Manned Penetration	6	8	2.10 (.99)	2.10 (.88)	.88
Finding Fixed Targets: NATO War, Restricted Visibility, Nonpenetration	7	9	2.00 (.94)	2.10 (.88)	.94
Finding Mobile Targets: NATO War, Restricted Visibility, Nonpenetration	8	7	2.00 (.94)	2.20 (.79)	.90
Finding Mobile Targets: NATO War, Clear Day, Nonpenetration	9	6	2.00 (.82)	2.30 (.82)	.83
Finding Fixed Targets: NATO War, Clear Day, Nonpenetration	10	11	1.90 (.74)	1.90 (.74)	.80
Finding Fixed Targets: Third Area, Restricted Visibility, Manned Penetration	11	17	1.80 (.92)	1.60 (.84)	.89
Coping With Missiles: Third Area, Clear Day, Manned Penetration	12	10	1.80 (.79)	2.00 (.67)	.85
Finding Fixed Targets: NATO War, Restricted Visibility, Manned Penetration	13	20	1.80 (.79)	1.60 (.70)	.85
Finding Fixed Targets: Third Area, Clear Day Manned Penetration	14	13	1.80 (.63)	1.80 (.63)	.72
Finding Fixed Targets: Third Area, Clear Day Nonpenetration	15	12	1.80 (.79)	1.90 (.57)	.70
Finding Fixed Targets: Third Area, Restricted Visibility, Nonpenetration	16	15	1.80 (.92)	1.70 (.82)	.94
Finding Mobile Targets: Third Area, Restricted Visibility, Nonpenetration	17	18	1.70 (.82)	1.60 (.70)	.54
Finding Mobile Targets: Third Area, Clear Day, Nonpenetration	18	19	1.70 (.82)	1.60 (.70)	.93
Coping With Guns: NATO War, Restricted Visibility, Manned Penetration	19	14	1.70 (.48)	1.80 (.63)	.87
Finding Fixed Targets: NATO War, Clear Day Manned Penetration	20	21	1.60 (.52)	1.50 (.53)	.82
Coping With Missiles: Third Area, Restricted Visibility, Manned Penetration	21	16	1.60 (.97)	1.70 (.95)	.95
Coping With Guns: Third Area, Restricted Visibility, Manned Penetration	22	22	1.50 (.71)	1.60 (.70)	.90
Coping With Guns: NATO War, Clear Day, Manned Penetration	23	23	1.50 (.97)	1.50 (.97)	1.00
Coping With Guns: Third Area, Clear Day, Manned Penetration	24	24	1.20 (1.03)	1.30 (1.06)	.96

The correlations for each item from round one to round two, as mentioned earlier, provide one form of test-retest item reliability. The mean of the observed reliability coefficients is .87 with a standard deviation of .11. Even with the very small sample of subjects used in this study, these results indicate a rather high level of test-retest item reliability.

The standard deviations (SD's) in Table 1 show an interesting trend from round one to round two. The median SD for the 24 items is .82 for the first round and .70 for the second round, a reduction of 15 percent from the first-round baseline. A nonparameteric sign test applied to the results in Table 1 showed 18 items with a lower SD in round two, four ties, and only two with a higher SD, which indicates that the convergence effect for the second round is significantly confirmed at the one percent level. This trend toward convergence of opinion in successive rounds is consistent with similar findings in the Delphi literature.

Figure 1 highlights the relatively small changes in mean item ratings from round one to round two. The x-axis represents mean first-round ratings, and the y-axis represents mean second-round ratings, with the importance rating scale spelled out for both axes. The circled numbers are the item numbers running from 1 to 24, the order of their appearance in the questionnaire, as shown in Exhibit A. Several data points are identical for two items; these are shown by placing item numbers nearbye with an "&" notation.

Figure 1 is basically a scatterplot of item means from round one to round two, roughly equivalent to the item rank-order correlation of .92 mentioned earlier. The diagonal line through the origin represents perfect correlation from round one to round two. Note that no average rating is less than 1, the point representing "marginal importance" for close air support R&D. Practically all items are closest to Step 2, representing an "important" R&D rating. This indicates that Giles Smith's item scheme for classifying close air support scenarios was fundamentally successful in tapping important mission scenarios as determined by panelist feedback.

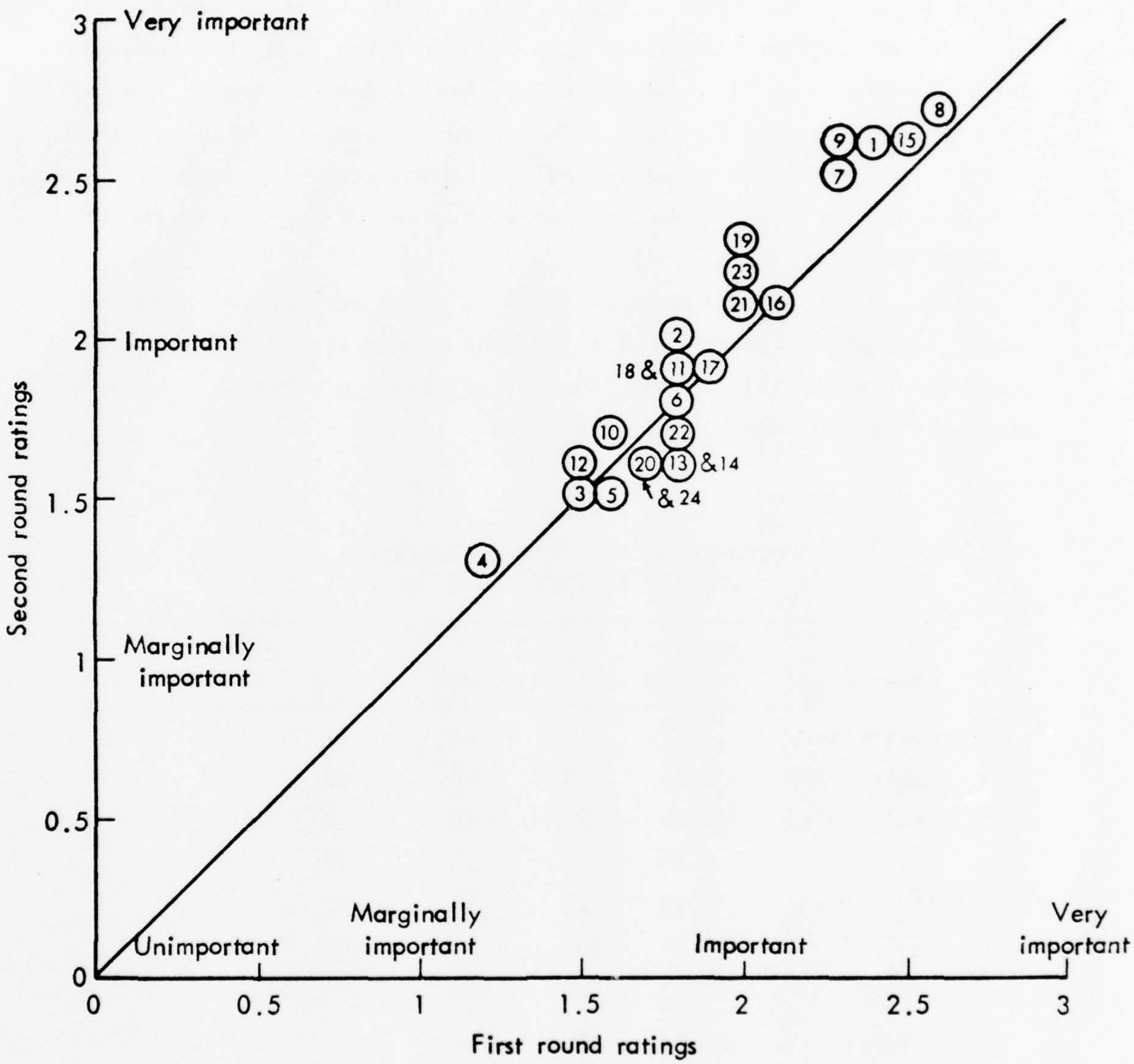


Fig. 1 — Mean ratings for first versus second round

4.2. Analysis of Variance for the Experimental Design

Table 2 shows the basic statistical design for participatory polling--a three-way analysis of variance for subjects, items and rounds. As mentioned earlier, there are three main effects, and three second-order interactions, with the three-way interaction serving as the error term to calculate the F-ratios. There are a total of 480 data points, or ratings for the two rounds combined (10 subjects x 24 items x 2 rounds = 480 ratings). Thus, the total degrees of freedom in the DF column adds up to N - 1 or 479.

Note, in the last column of Table 2, under statistical significance, that mean differences for four out of six tested effects are statistically reliable at the five percent level or better. These are discussed in the order shown in Table 2.

Table 2

THREE-WAY ANALYSIS OF VARIANCE:
ROUNDS, SUBJECTS, AND ITEMS

Source	Sum of Squares	DF	Square	F-Test	Signif.
Rounds(R)	.41	1	.41	4.91	.05
Subjects(S)	51.16	9	5.68	68.48	.01
Items(I)	64.09	23	2.79	33.58	.01
RS	1.18	9	.13	1.58	NS
RI	2.29	23	.10	1.20	NS
SI	190.74	207	.92	11.09	.01
RSI	17.13	207	.08	-	-
Total	326.99	479			

Although the mean importance rating difference is significant at the five percent level for round one versus round two, the actual difference is not substantial. The mean rating for round one is 1.89 as compared to 1.95 for round two. This very slight difference is probably not interpretively important. Observed subject differences are more substantial.

The expert with the lowest mean rating was at 1.3 whereas the highest subject mean was at 2.3. Freely translated into the importance scale, experts varied in their overall mean assessment of the close air support mission from "marginal" to "important."

Item differences were also substantial as well as statistically reliable, ranging from a mean rating of 1.65 for the least important item, to 2.65 for the most important item. The pattern of item differences is illustrated in Fig. 2 which was conceived and prepared by Giles Smith. This figure shows the full rating scale on the y-axis, and lists all 24 items on the x-axis according to the classification scheme for close air support scenarios. The "bullet" data points represent the 24 item means for second-round results.

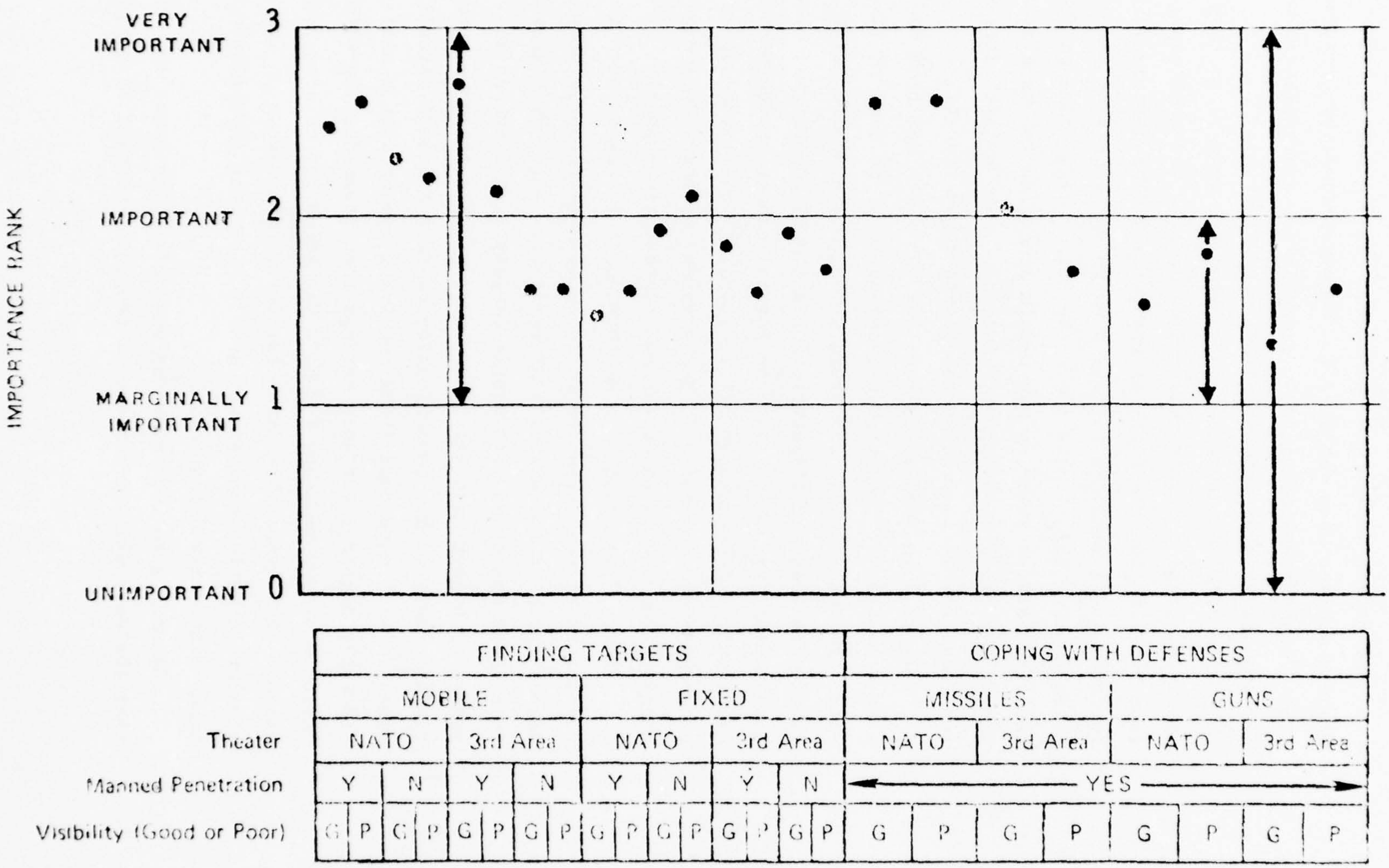
Figure 2 is read as follows for the fifth dot from the y-axis with a mean at 2.7 and arrows going up to 3 and down to 1. The item refers to the scenario concerned with finding mobile targets (top row), in a third area war (theater), for a manned penetration mission, with good visibility. Note that the logical scheme at the bottom of Fig. 2 completely differentiates all 24 items. This is the top-rated item with a second-round mean of 2.7, and a spread of individual expert ratings ranging from one to three (arrows representing range).

Figure 2 is brought in at this point in the results to illustrate the nature of the large and reliable mean differences between items demonstrated in the analysis of variance. Note, by visual inspection, that items concerned with mobile targets (e.g., troops and tanks) have generally higher mean importance ratings than items concerned with fixed targets (e.g., communication centers and anti-aircraft installations). Also note that items involving coping with missiles have considerably higher importance ratings than items dealing with guns. There is also a tendency for NATO war scenarios to show higher mean ratings than comparable items for Third Area wars. On the other hand, similar scenarios contrasting good versus poor visibility conditions tend to show mixed results.

Returning to the three-way analysis of variance in Table 2, note that the only significant second-order interaction is subjects versus

Fig. 2

RESULTS OF EXERCISE: RANKING OF CLOSE AIR SUPPORT OPERATIONAL CAPABILITY OBJECTIVES



items. The other second-order interactions between subjects and rounds, and items and rounds are statistically nonsignificant, largely because of the general stability of ratings noted earlier from round one to round two.

The 10 subjects versus 24 items interaction refers to the variability of the 240 means for this second-order effect. This significant interaction indicates that the unique combination of a particular subject rating a particular item is different from the other subject-item combinations, after all other first and second-order effects are subtracted out. Thus, after accounting for individual expert differences, item differences, and rounds, the rating given by an expert to a particular item shows stable and reliable differences among the 240 combinations investigated in this study. This result has special significance for the dynamics of expert opinion that are discussed in the interpretation section.

Table 3 shows the analysis of variance results for panelist reasons listed for the 24 questionnaire items in the second round. Recall that each item in the second round was accompanied by three to five reasons derived from first-round feedback, and that these reasons subsequently received importance ratings in the second round in arriving at final overall importance ratings for the given item. Thus, each item constituted a data matrix of 10 subjects versus three, four, or five reasons with a subject-reason rating in each of the cells. This two-way matrix permitted an analysis of variance for subjects versus reasons for each of the 24 items.

The analysis of variance is split into three components: 1) subject differences (in rating reasons); 2) differences in mean ratings for the listed reasons; and 3) interaction variance which is used as the error term, since there is only one observation per cell. Table 3 lists subject differences and reason differences for each item, indicating whether each separate analysis of variance found such differences to be statistically significant or nonsignificant. A "dash" entry in Table 3 means nonsignificant, in contrast to the indicated five percent and one percent levels. Exhibit B should be consulted for the specification of the item scenario matched to the item number

Table 3

COMPOSITE ANALYSIS OF VARIANCE RESULTS FOR ITEM REASONS

Item	Subject Differences	Reason Differences	Item	Subject Differences	Reason Differences
1	-	.01	13	-	-
2	-	-	14	-	-
3	-	-	15	-	.01
4	-	-	16	-	-
5	-	-	17	-	-
6	-	-	18	-	-
7	-	.05	19	-	.05
8	-	.01	20	-	-
9	-	.05	21	-	-
10	-	-	22	-	-
11	-	.01	23	-	-
12	.01	-	24	-	-

shown in Table 3, and for listed reasons and mean ratings obtained for each reason from round two.

There are two notable results in Table 3. The first is that differences between subjects in rating reasons are basically nonsignificant. Only one item (number 12) shows statistically reliable differences between experts in ratings for reasons. One item out of 24 would be expected on the basis of chance alone, assuming independence of items. Thus, although experts differ significantly in importance ratings between items as shown earlier in the three-way analysis of variance, they do not differ significantly in evaluating reasons within the setting of a particular item.

The second notable result in Table 3 is that seven out of 24 items show reliable mean rating differences between reasons listed within items. Only one statistically significant item out of 24 would occur on the basis of chance at the five percent level, assuming independence of items. This result indicates that experts can reliably discriminate more important from less important reasons in evaluating item scenarios.

4.3. Qualitative Results

Table 4 is virtually a verbatim reproduction of open-end subject responses to the item concerning "insight" in the overall evaluation form: "What was your most important insight in working with this questionnaire?" This complete set of responses for the 10 participants shows the range of reaction to the questionnaire procedure, illustrating strengths and weaknesses. The reaction of the expert panel to the overall evaluation form is summarized on the last page of Exhibit B, and is discussed in the interpretation section which follows.

Table 4

REPORTED INSIGHT

1. Spread of "reasons" for each topic suggests that paper summarizing (1) user's view (i.e., the operating unit), (2) the designer's view, (3) the historian's view, (4) the scientist's view is practically mandatory.

2. Risk and futility of manned penetration.

3. Need more explicit explanations of reasons for importance. Without good reasons, I find myself sticking to my previous choices, but without defending them well.

4. If every respondent understood the questionnaire, there sure is a wide variation in response; in fact, some of the responses seemed nonsequitur.

5. The diversity of opinion (and the supporting reasons) with regard to manned a/c penetration.

6. None gained.

7. Importance of clear, unambiguous statement of questions and answers.

8. Appreciate divergence of views about "CAS" mission concept. But really, I've learned little new.

9. Stereotyped views of warfare: ability to separate targets/defenses.

10. Developing relative priorities on targets and force allocation.

To obtain the best picture of the great variety of reasons offered for and against various types of mission objectives, the reader should consult Appendix B which lists second-round results. The spread of open-end qualitative reasons is tabulated for each questionnaire item, illustrating the potential of participatory polling for eliciting and explicating adversary positions.

5. INTERPRETATION

The interpretation section covers four areas: methodological problems, the analysis of expert opinion, evaluation of participatory polling, and follow-on recommendations. The methodological interpretation includes experimental procedures and questionnaire technique; the analysis of expert opinion covers similarities and differences between experts, and the psychology of expert opinion; the evaluation of participatory polling compares observed results against the eight key objectives for this technique; and the recommendations for follow-on work cover various extensions of the method that seem promising for future research.

5.1. Issues in Experimental Design

It was stated at the outset that this was an informal exploratory study to gain some insight into the feasibility, acceptability, credibility, and usefulness of participatory polling in a long-range Air Force planning application. This investigation clearly established the feasibility of participatory polling. There were no inordinate tactical, time or cost problems encountered in implementing the technique compared to ordinary problems in designing and using opinion questionnaires.

Developing a suitable questionnaire covering operational requirements for the close air support mission was indeed a difficult task, but one that we feel was handled with reasonable success within available resources. Content validity was comprehensively covered and systematically sampled by the five-way classification scheme developed by Giles Smith to generate the 24 items in the questionnaire. The strength of this scheme lay in the interaction effects among the five scenario variables specified by individual questionnaire items. Its major weakness, as pointed out by some of the panelists, was in not being able to distinguish separate effects of each variable in isolation from the other variables.

The problem of selecting experts according to explicit sampling criteria was not formally addressed in this exploratory study. Expert colleagues were personally solicited from a highly sophisticated group

of military analysts at Rand; practically all agreed to see the study through; and there were no dropouts from round one to round two. Accordingly this limited study does not have much to say about the rather important problem of sampling diverse experts at remote locations, and matching their skills and background against the planning task at hand. The use of an open meeting among the panelists to explain the study and the questionnaire is consistent with the Rand professional "culture," and probably helped to enhance competition, interest, and motivation.

Perhaps the most unique characteristic of participatory polling is the feature requiring written justification of ratings in the first round. Were there any special difficulties encountered in obtaining and processing these open-end verbal responses? We were not completely successful in getting unique "reasons" for each item rating. Approximately half of all these reasons for all panelists combined were "dittoes" or repetitions of reasons offered earlier. In other words, for the typical first-round questionnaire, about half of the items would have unique and different reasons associated with ratings, and the other half would show direct replications of earlier reasons, often with ditto marks. This "ditto phenomenon" is clearly one of the key procedural problems of participatory polling. In the ideal situation, 100 percent of all responses would be unique.

In our case, the 50 percent ditto rate is attributable largely to the considerable content overlap between items. For example, two items concerned with coping against missiles in a NATO war for a piloted aircraft mission would differ only in clear versus restricted visibility. If the respondent did not believe that weather conditions were critical, he would be likely to offer the same rating and give the same justification for both of these items. The same tendency would apply to items sharing the same one or two salient features that the panelist believed to be critical in his ratings. Accordingly, the relatively high "ditto rate" in this study is largely attributable to correlated, overlapping items generated by the item classification scheme. Questionnaires based on relatively independent items are probably less likely to evoke similar or identical verbal responses.

Once the first-round reasons were collected, the task was to sort and classify them for second-round presentation and evaluation by the expert panel. Computerization of verbal responses helped considerably in facilitating this task by simply making it possible to collate verbal responses for each questionnaire item on a separate "display" printout. This made visual inspection much easier for grouping and comparing responses for the second round.

The transition from raw verbal data in the first round, to processed "reasons" for the second round, was based on subjective evaluation by the author following six rules or guidelines:

1. Combine similar or equivalent reasons.
2. Stick closely to the original language, retaining the flavor of the polemics (e.g., "stop tanks or lose war.").
3. Eliminate expletives that merely repeat the item without offering substantive justification for the subject's rating (e.g., "This is a prime job for the close air support mission".).
4. Work toward an adversary balance across all reported views.
5. Leave no distinguishable reasons out, unless they are essentially covered by similar reasons.
6. Randomize the order of reasons for successive items so that the same individuals and their viewpoints are not consistently represented at the beginning, middle or end of the listed reasons for each item.

The above six guidelines were not difficult to follow for 10 subjects which resulted in three to five reasons associated with each item in the second round. If many more respondents were used, more selective techniques would probably be necessary to keep the number of reasons per item within manageable size for the second round. A useful suggestion from the panelists is that "pro" reasons appear in one grouping with "con" reasons in another grouping for each item to heighten or polarize the adversary effect. This seems like a worthwhile suggestion for situations where there are sufficiently large

samples of reasons to develop detailed "pro" and "con" arguments in a balanced manner. In any case, the systematic feedback of thoughtful adversary reasons in participatory polling was reported as the single most valuable feature in generating the main insights achieved by participants in this exercise. This conclusion was verified by the responses of the panel to the overall questionnaire evaluation in the second round (see last page of Exhibit B).

The second set of considerations on experimental methodology is concerned with questionnaire reliability and validity. Data on both stem from the various types of item analyses that were performed.

One approach to item reliability was test-retest correlations for the first versus the second round. The mean observed test-retest correlation was .87 with a standard deviation of .11 for the 24 item coefficients. This is quite high and is probably an overestimate in certain respects. In round two, respondents had copies of their first-round responses; this situation, peculiar to this method, does not usually hold for conventional test-retest reliabilities. With first-round "answers" in front of him, it is easier for the respondent to give identical or similar second-round ratings.

Another approach to test reliability was obtained in the second-round "Overall Evaluation" questionnaire where the panelist was asked to rate the overall reliability of his estimates and responses to the questionnaire. Only one subject felt that his responses were "marginally reliable"; all others rated their responses as "fairly reliable" or "very reliable" (see Exhibit B, last page).

Still another approach to test reliability are the analyses of variance conducted on item ratings and item reasons reported in the results section. Recall that the three main effects were found to be statistically reliable in the three-way analysis of variance (subjects, items, and rounds), and that the panelists were able to distinguish more important from less important reasons for items when the ratings for reasons were subjected to analyses of variance. Thus, the three key variables in the basic experimental design, and the reasons for ratings, were able to show reliable statistical effects even with the

small samples of subjects and items used in this study. The test-retest, subjective evaluation, and analysis of variance approaches to questionnaire and item reliability concur in showing a generally high and useful level of test-reliability in this application of particpatory polling.

One important type of procedural reliability for participatory polling was not tested in this study. This is inter-rater reliability in screening and filtering first-round verbal responses for second-round presentation. Given a set of ground rules such as the six guidelines mentioned earlier, two or more judges or raters should independently produce item "reasons" for second-round feedback. Comparison of such independent listings of item reasons would provide the basis for inter-rater reliability in generating second-round feedback from the first-round protocols. Due to time and resource constraints, this type of reliability was not tested in this study.

There are four basic types of validity associated with questionnaires: content, construct, predictive, and concurrent validity. These are briefly discussed to round out this section on methodological interpretation. Content validity was mentioned earlier in connection with the five-category classification scheme for close air support scenarios as used in this study. Construct validity pertains to the concepts and theories underlying the participatory polling technique and the close air support questionnaire for long-range R&D Air Force planning as used in this investigation.

Two types of validity remain to be discussed, predictive and concurrent. Predictive validity refers to the effectiveness of the questionnaire in predicting specified criterion events related to performance effectiveness in the area studied. This may take a variety of forms in the Air Force planning context, such as technological forecasting, military budget predictions and forecasting international crises. None of these has been incorporated within the limited scope of this demonstration prototype of participatory polling. This does not mean, of course, that the criterion prediction problem is of minor consequence. In this study, participatory polling is aimed at supporting the Air

Force policy-maker by providing him with a suggestive range of expert opinion and associated adversary rationale for R&D planning in a limited area. Predictive criteria could have been worked into the study design, but not without greatly increasing the scope of the investigation.

Concurrent validity refers to the logical internal consistency of the technique, the questionnaire, and the various performance measures used. For example, are quantitative ratings and open-end responses logically consistent with each other? Have the various objectives of participatory polling been met in this initial demonstration? Are the illustrative military opinions and findings credible for planning purposes? The analysis of such outcomes, comprising various approaches to concurrent validity, represents the bulk of the interpretation that follows in the analysis of expert opinion and participatory polling.

5.2. Analysis of Expert Opinion

The general intent in this section is to review the empirical findings on expert opinion in this study, and then to briefly extrapolate the potential of such findings to behavioral analysis of expert opinion more broadly considered. The current findings are examined from the viewpoint of similarities and differences between experts, whereas the behavioral analysis examines differences between experts and nonexperts.

Perhaps the most striking feature of the results are the quantitative and qualitative differences within this relatively small sample of experts working in virtually the same professional environment at Rand. In reporting their main insight from this study, the participants use such terms as "diversity of opinion," "divergence of views" and "wide variation in response" (see Table 4). This appreciation of the great range of expert opinion seems to suggest that perhaps the distinguishing characteristic of experts lies in well-defined differences in values, assumptions, approaches, and in personal judgments reflected in current controversies in the domain of inquiry. This interpretation supports a basic premise and objective of participatory polling--the proper role of experts is to map out and explicate the

range of alternatives in the domain of inquiry in a balanced adversary setting. The focus is on expert differences rather than expert consensus, on widening the range of thoughtful and explicated alternatives and policy choices as opposed to premature closure in contested areas.

There are two key statistical findings supporting this emphasis on the differences between experts, both from the three-way analysis of variance reported in the results (Table 2). The first is large observed differences between subjects, that is, the mean ratings for the 10 subjects varied considerably. What are some of the explanations for this statistically reliable result? One relatively straightforward explanation is that some experts saw the close air support mission as more important for R&D planning, whereas others saw it as less important. This type of halo effect is undoubtedly true to some extent, depending upon personal interests and preferences.

The observed differences, however, are partly contaminated by unavoidable limitations of the polling technique. Some of these differences may be due to different subjective "anchoring points" in using the four-step scale; that is, some participants may have felt more comfortable using the upper or lower parts of the importance scale in terms of their individual interpretation of the four steps. This "anchoring" phenomenon is a type of an estimation halo effect that cannot be parceled out of the observed results with available data. In retrospect, the questionnaire should have included some "control" items asking respondents to rate their estimate of the general importance of the close air support mission against competing Air Force missions; it is hypothesized that a positive correlation would occur between mean close air support item ratings and the overall comparative close air support rating, providing some indication of the strength of the aforementioned general mission halo effect in this study.

The second key finding in the three-way analysis of variance, in this context, is in the significant interaction between subjects and items. This result raises some key issues on the nature of overall expert differences. That is, over and above the separate effects of

subject, item and round differences, the panelists showed unique and stable preferences for some mission scenarios as compared to others. These preferences were unique for particular individual-item combinations. This result suggests that, in addition to overall value differences within the domain of inquiry, there also exists a highly differentiated superstructure of individualized preferences that are stable (not changing from round to round). In this study, the 10 subjects versus 24 scenarios resulted in 240 subject-item combinations that comprise the "differentiated superstructure."

The observed significant interaction between experts and specialized issues suggests a differentiation hypothesis for distinguishing between experts and nonexperts. That is, experts are more likely to exhibit a uniquely specialized and differentiated set of firmly held opinions in their area of expertise. In contrast, nonexperts are hypothesized to have less specialized opinions, be more suggestible, and be more likely to change their opinion with group feedback. This "differentiation hypothesis" stems from learning theory, and presupposes a generalization/specialization progression as the individual gains increasing experience and makes progressively finer discriminations in his area of expertise. The empirical test of this hypothesis would be comparative experimentation between experts and nonexperts, particuarly in the extent and significance of the subject-item interaction, for the three-way analysis of variance paradigm of participatory polling (subjects, items, and rounds).

The observed similarities among experts are also instructive, in contrast to observed differences between experts. Two statistical results indicated similarity among experts. One was the finding that the subjects versus rounds interaction was not statistically significant in the three-way analysis of variance. Only 16 percent of all ratings were changed from the first to the second round. This result underscores the relative stability of expert opinion mentioned as part of the differentiation hypothesis.

The second statistical example of similarities between experts was the absence of significant individual differences in rating the

importance of reasons put forth for each item in the second round. In the two-way analysis of variance results for subjects versus reasons, only one item out of 24 showed statistically significant subject differences (Table 3), a result to be expected on the basis of chance. In contrast, the experts were often able to reliably discriminate more important from less important reasons within each item, as shown in Table 3. These two results from Table 3 suggest that experts do show consistent consensus when judgment is exercised in a relatively well-defined and highly specialized area.

The judgmental setting of comparative reasons for specific item scenarios was sufficiently specialized to lead to expert concurrence. These results suggest that the more highly specialized the opinion, the more likely are experts to concur with each other. This expert concurrence hypothesis is based on the assumption that experts have a common data base of basic facts, and extensive agreement on lower-level or specialized opinion stemming from factual or descriptive data. In contrast, nonexperts, lacking an extensive common data base, are hypothesized to show more of a random, nondiscriminating pattern in evaluating reasons for adversary positions. In the present setting, for example, it would be hypothesized that nonexperts would not be able to reliably rate some reasons as consistently more important than others for the various items.

The differentiation hypothesis and the concurrence hypothesis map out a suggestive relationship between expert opinion and the level of generality of the domain of inquiry. For more highly specialized areas, experts are more likely to concur; for more generalized areas, experts are more likely to disagree, but within a relatively consistent and structured framework of opinions. For nonexperts, it is hypothesized (without any direct evidence from this study) that they would show less concurrence in specialized opinion, and a more stereotyped range of differences in more general opinion. The basic mechanism at both ends of the specificity-generality continuum is that experts possess finer discrimination capabilities based on larger and relatively well-structured repertoires of facts and opinions, and that they are

capable of organizing such repertoires in more novel combinations at more general levels of opinion.

The main challenge in conducting an expert opinion study is maximizing the available expertise in relation to study objectives. This leads to the "snap judgment" hypothesis: as human information processing time approaches zero, the quickest possible snap judgment reaction time, so does the difference between expert and nonexpert opinion tend to disappear into common baseline stereotypes. The snap judgment hypothesis basically asserts that considered opinion is more likely to enhance differences between experts and nonexperts than snap judgments. In this study, the participatory polling technique was deliberately designed to minimize snap judgments in item ratings and to force a modicum of reflective thought. In the first round, the requirement to specify a reason for each rating probably forced participants to consider more alternative judgments for each item; and in the second round, the requirement to rate three to five reasons for each item was a way of forcing participants to recognize various aspects of the scenario before assigning their final rating. Conventional questionnaires invite quick-scan, quick-response information processing that often approaches the snap judgment extreme.

Basic research is required for the various forms of behavioral activity subsumed under the general rubric of expert information processing. At one extreme lies the different varieties of snap judgments made almost instantly without reflection. These may be valuable in revealing stereotypes or as projective techniques to explore personality characteristics associated with particular types of opinion. Snap judgments probably offer more information on the subject than it does on the attributes being investigated. At the other extreme are techniques that involve extensive preparation and work on the part of the expert before he submits his considered opinion. This end of the continuum overlaps with combinations of opinion with formal procedures, particularly where opinion is characterized by extensive qualifications, proofs, and arguments of the position taken. Participatory polling

is one example of deliberate questionnaire design to force panelists into structured reflection and analysis before responding to each item.

5.3. Evaluation of Participatory Polling

This section evaluates the effectiveness of participatory polling, as used in this initial demonstration, in meeting the eight objectives of the technique described in the introduction.

5.3.1. Expert-Interpreted Opinion. The technique in the present setting was moderately successful in eliciting expert-interpreted opinion. The 50 percent "ditto" rate mentioned earlier, where only about half of all items were accompanied by unique reasons, is largely due to overlapping, part-whole items stemming from the classification scheme used for mission scenarios. The "ditto" problem can probably be minimized by using relatively independent questionnaire items with minimal overlap in content. In any case, the reasons that were solicited represented the basis for the main insights reported by participants at the end of the second round. These reasons also provided direct interpretations of expert opinion in a systematic manner across the entire sample of panelists, as opposed to more indirect interpretations by investigators with conventional questionnaires.

5.3.2. Balanced Adversary Procedure. An approximation to a balanced adversary procedure was achieved in presenting essentially all distinct reasons for each item in the first round for review and evaluation in the second round. A larger number of subjects, perhaps 20 as opposed to the 10 used in this study, would have provided a larger number of reasons offered in support of first-round opinion. A larger pool of reasons, in turn, would permit more formal classification of reasons into "pro" and "con" positions to heighten the adversary effect as suggested by some panelists. The adversary procedure was "balanced" in two key respects in this demonstration. First, all subjects were required to provide reasons for their first-round ratings, as opposed to voluntary submission. Second, the procedure treated all individuals equally whether they were outliers or conformists.

5.3.3. Shared Knowledge of Results. A key intent in participatory polling is to make available to all members of the panel the knowledge gained from first-round results to improve second-round judgments. In this study, panelists had the following information for their final second-round ratings: first-round ratings for the group, including means, standard deviations, and frequency distributions, and the full range of reasons offered for first-round opinions. To the extent desired, they were able to use any or all of these data for their final evaluations. This exercise was only moderately successful in supplying optimal knowledge of results. Several panelists complained that the reasons were not spelled out in sufficient detail to be convincing. This demonstration did not systematically test the impact of more detailed justification of ratings. The open-end comments indicate that a greater level of detail in questionnaire items and in panelist reaction would probably have made a greater impact on second-round opinion.

5.3.4. Group Data Base. Some of the preceding comments apply to the adequacy of the group data base. A larger number of subjects would have generated more extensive adversary arguments in contested areas. A more detailed set of reasons (e.g., a paragraph rather than a sentence for each item) would probably have produced more compelling reasons to modify ratings.

The group data base, at the end of two rounds, did meet certain practical requirements. Mean item ratings were reliably different from each other--the missile threat was seen as more critical than the gun threat, mobile targets such as troops and tanks were more important than fixed targets--nonsurprising results. Differences between items were reliable and test-retest reliability was high for quantitative measures. The composite quantitative and qualitative group data base for this study offers suggestive preliminary data illustrating Air Force mission requirements for close air support R&D planning.

5.3.5. Participatory Evaluation. The panelists provided considerable feedback on items in particular, rating procedures, and the test as a whole. Virtually nothing escaped their critical evaluation. It

is impossible in a report of this limited scope to document all the comments received. Nonetheless, the instructions, procedures, forms, and spirit of this technique encouraged panelists to act as experimenters as well as subjects, and the study benefitted considerably from their participation. The participatory objective was successfully met in this initial study.

5.3.6. Flexible Anonymity. The participants had no complaints concerning the level of anonymity used in this study. The initial group indoctrination session allowed panelists to know who participated in the study. Individual responses to the first and second-round questionnaires were kept anonymous, by mutual consent and prior agreement. This turned out to be quite desirable because of the heat generated by some of the more controversial issues (e.g., preference for manned versus unmanned aircraft for future R&D, see Exhibit B). Finally, in this final report, the responses and comments of individual panelists have also been kept anonymous. There is nothing sacrosanct about the particular manner in which anonymity was handled in this study. Other inquiries may involve other approaches to anonymity, depending upon study objectives, hence, the desirability of flexible anonymity based on mutual consent, specified in advance. The objective of flexible anonymity was successfully met in this initial demonstration.

5.3.7. Scientific Standards. This demonstration was carried out as an informal exercise as specified at the outset. However, within the limits of the informal exercise, a variety of experimentally-oriented procedures and statistical tests were explored to see how they would fare in the participatory polling setting. Efforts were made to assess the reliability and validity of the procedure including extensive statistical tests for various aspects of item reliability. Unexplored scientific areas were pointed up where no effort was made and where no tests were conducted (e.g., expert sampling parameters, predictive validity, inter-rater reliability in processing panelist reasons, and testing the limits for more detailed items and reasons). The tests that were conducted, however, provide a useful framework for more scientific studies (e.g., the basic three-way analysis of

variance experimental design for participatory polling, testing mean differences between importance ratings for item reasons, various measures of test reliability, and useful psychometric feedback on scaling and rating procedures).

5.3.8. Cost-Effective Technique. The time, effort and skills required for participatory polling are roughly equivalent to those required for designing and developing comparable opinion questionnaires. There is a significant amount of additional work required in processing first-round reasons for second-round presentation. In this case, it amounted to less than a man-week of effort, which is not at all prohibitive. The rich data yield from adversary reasons are easily worth this amount of extra effort. Computerization of all quantitative and qualitative data greatly facilitates preparation of second-round materials as well as statistical testing and analysis of the results. For large samples of open-end data, computer-aided analysis of verbal data holds considerable promise and is already well advanced in the social sciences. Participatory polling has valuable positive fallout in facilitating the interpretation of the results for the investigator. Comments and recommendations of panelists help immeasurably and add diversified viewpoints in analysis of the data and in preparation of the final report. All in all, this initial trial of participatory polling has demonstrated great potential as a cost-effective technique for soliciting credible expert opinion.

6. RESEARCH RECOMMENDATIONS

There are four key areas for follow-on research derived from the current exploratory study. All recommendations reflect the personal opinion of the author, and do not necessarily reflect Rand opinion. The first is an extension of the Air Force R&D planning base. The second is conducting a more formal experimental study in participatory polling. The third is computerized analyses of open-end responses, and the last is basic research on expert opinion.

6.1. Conclusions from the Experiment

The current demonstration study is strictly illustrative, and is concerned only with the close air support mission. There are about a dozen other recognized Air Force mission areas in long-range R&D planning such as strategic offense, air defense, interdiction, command and control, etc. Follow-on studies could focus on individual areas in stepwise fashion to build up an inductive data base, or they could, as recommended by some of the panelists, provide comparative ratings for all Air Force missions as is the case with periodic budget review. The latter approach would provide a deductive data base. There are advantages and disadvantages to both approaches which are not discussed here since the intent is primarily to point to possible alternatives for future work.

Another direction for these studies is to solicit expert opinion in relation to competing candidate projects, such as new weapons, new communications technology, and improved management systems. Our original concept was to obtain expert opinion for close air support mission requirements and then, with baseline mission requirements established, proceed to rate competing candidate Air Force projects on their cost-effectiveness in achieving desired mission objectives. This approach could also be employed stepwise, on a mission-by-mission basis, or it could be implemented across all mission areas as described above. With either approach, the end result would be a systematic set of recommendations, with extensively evaluated reasons for such recommendations, which would include 1) preferred mission requirements, and 2) preferred candidate R&D projects best meeting

such requirements. These would provide a ranked set of alternatives for policy-makers to examine in determining relative priorities for long-range R&D planning in the Air Force.

A notable caveat should be mentioned for the above approaches. The expert panels should not consist entirely of one particular interest group, such as Rand analysts as used in the current demonstration. Various expert groupings should be tested for more balanced adversary results (e.g., operating command personnel, design and development specialists, senior policy-makers, and scientist-analysts). This recommendation came from our panelists and is supported by the research literature which shows clustering of opinion based on background (e.g., Derian and Morize, 1973). This type of safeguard is essential for controversial areas with major vested interests.

6.2. Formal Experimental Studies in Participatory Polling

The results of the present demonstration study have been sufficiently encouraging to recommend *a formal experimental test of participatory polling* capitalizing on the prototype experience. The formal experiment should be more comprehensive than the prototype in a variety of ways, and should provide more definitive tests of the value of participatory polling.

In the area of expert selection, one extension has just been mentioned above. Better checks and balances in vested expert opinion are obtained in the Air Force setting by selecting experts with diverse backgrounds. Such selection should match explicitly specified skills against the technical content of the questionnaire, and such matching can be verified by adequate data on expert sampling parameters. This information can be easily obtained by a biographical data questionnaire.

The expert panel should include at least 20 participants to ensure a sufficient diversity of adversary reasons for each item and to provide adequate sample sizes for the various statistical procedures employed. Panelists could respond to self-administered questionnaires after an initial briefing, as in the present study, or a structured,

formal interviewing technique could be used, particularly for longer questionnaires more likely to induce respondent fatigue or lower his motivation.

The methodology of the formal study should take full advantage of the lessons learned with the prototype. Rating scales could have additional steps to accommodate the full dispersion of opinion. Item design could take on a variety of forms quite different from that used in this study. Items should be sufficiently distinct and nonoverlapping to minimize "ditto" justifications. Item scenarios, in Air Force settings, should probably be more detailed, and panelist responses in justification of item ratings might also be more detailed, perhaps up to a paragraph in length as opposed to single sentence size. Item and open-end response detail could, in fact, be independent experimental variables by using two contrasting levels of detail.

Open-end responses to the first round should be processed by two independent raters following a common set of guidelines to arrive at second-round inputs. This would provide the basis for more objective determination and presentation of adversary reasons and for determining inter-rater reliability which was not done in this prototype study. The second-round listing of item reasons could be presented in "pro" and "con" adversary format if the content area and data permit balanced presentation.

The experimental design should permit analysis of variance tests for differences between experts, items, rounds and reasons along the general lines followed in the current study. The various types of reliability and validity discussed in this study should be tested in the framework of a systematic item analysis.

Questionnaire forms should be subjected to preliminary pilot testing to debug errors and weed out poor items. The quantitative and qualitative data bases should be computerized, as in the present study, to facilitate processing and analysis. Issues connected with computerized analyses of open-end responses are discussed in the next section. If a follow-on formal experiment follows these various guidelines it is likely that participatory polling would receive a systematic, definitive test of its assets and liabilities.

6.3. Computer-Aided Analysis of Open-End Responses

A special characteristic of participatory polling is the systematic use of open-end responses to establish an adversary data base. This recommendation explores some of the types of computerized research that are needed to facilitate more cost-effective processing of such problematic responses.

The open-end item, characterized by free verbal responses, poses special problems, not only for participatory polling, but for questionnaire design broadly considered. Such items are used for a variety of reasons. One is to obtain unique identifiers or basically qualitative information; for example, an individual's name and occupation. Another is to build an empirical data base inductively in a new or unexplored area; for example, organizational and professional affiliations of individuals, reasons for expert opinions as in participatory polling, or spontaneous evaluative comments over and above those listed for structured questions. The temptation is often great to include numerous open-end questions, particularly in unstructured areas of inquiry.

The price can be very high for such inclusion because of the imposed burden of extra interviewing time or respondent time where responses are recorded verbatim rather than merely checked off, because of greater data storage requirements for verbal as opposed to quantitative data, and because of the considerably greater amount of time required by the researcher to organize, classify, and analyze open-end responses. Investigators with tight schedules often ignore open-end data gathered at great time and cost, or perhaps eyeball the responses in a cursory manner for global impressions, or simply give up after an abortive attempt to investigate such responses systematically when they find out that it involves more time and effort than they anticipated. The tendency is to look the other way, to continue gathering costly open-end data that will not be analyzed, and hope that someone, somehow, might get around to it at some propitious time. Open-end items are generally more costly than objective items from the point of view of data collection, reduction, and analysis. They are prime

candidates for being dropped to make room for useful and important items that will be analyzed and will in fact contribute to the study.

How should open-end items be handled? The basic recommendation is that they should compete with objective items and with each other on the grounds of item reliability, item validity, research contribution and cost. This means that open-end items--such as those used in participatory polling--should undergo some form of item analysis to determine their value.

But what does item analysis mean for open-end responses? Ten basic steps may be identified as shown in Table 5 which indicates potential use of computers for each stage. The first step is "cleaning" and editing responses for clarity and removal of redundancy. The second step is to keypunch verbal responses and store them in computer-accessible form.

The third is to provide verbatim printouts of item responses classified in some convenient manner for initial inspection by investigators. This stage is equivalent to the initial listing of marginal frequencies for objective items. With computer-accessible responses, such printouts can be provided on a selective basis by simple cross-tabulation; for example, listing by age or sex group, educational level, professional specialty, geographical area, or other categories, according to the investigator's interest and leading hypotheses. Such listings would help partition large samples into more manageable subsamples. If appropriate software is available, concordance listings may be provided of key words or phrases; for example, a simple KWIC listing (Key Word In Context) may provide the equivalent of a crude frequency count for more "popular" responses.

The fourth step is to analyze such listings and determine whether further formal analysis is desired. If so, a rating scheme may be designed by the investigator yielding quantitative scores that would permit intercorrelation or other statistical analyses with respect to other open-end items or with objective items.

The fifth step is development of rules and standards for rating scales. A simple example would be a bivariate classification such as economic reasons versus other (noneconomic) reasons for an open-end

Table 5

BASIC STEPS IN COMPUTER-AIDED ANALYSES OF OPEN-END VERBAL RESPONSES

	Step	Possible Use of Computer
1.	Clean and Edit	Partial
2.	Keypunch	Yes
3.	List Responses	Yes
4.	Examine Responses	Yes
5.	Develop Rating Scale or Classification Scheme	Partial
6.	Rate Responses Independently	Partial
7.	Compare Independent Ratings	Yes
8.	Adjudicate Major Differences	Partial
9.	Analyze Ratings	Yes
10.	Incorporate Scale into Instrument	No

item concerned with competing R&D projects. Such a scale would permit bivariate intercorrelations with other items and would lend itself to other types of statistical analyses. There are many other types of scales, such as comparative importance ratings of reasons for opinions as used in this exercise with participatory polling.

The sixth step is independent application of the rating scale to individual responses by at least two different raters. Computer assistance may be possible in aggregating responses for easier scoring at this step. The seventh step is comparison of independent ratings (e.g., inter-rater reliability or "conspect reliability," expressed as percentage of exact agreement or as a bivariate correlation coefficient). This procedure was recommended earlier in this study to evaluate the reliability of reasons selected for second-round feedback.

The seventh step provides hard-copy feedback on item reliability. The eight step is adjudication of substantive differences in independent ratings, either by consultation of the raters or by an external

"judge." The ninth step is analysis of the scaled values of the open-end item against other items of interest in the overall instrument. This step provides feedback on concurrent item validity. The tenth step is possible transformation of the open-end item to an objective item incorporating the new scale in the next application of the instrument. These 10 steps apply to almost any type of open-end item format, and have generality beyond participatory polling. Computer-aided research, perhaps using current technology in natural-language processing, could lead to significant advances in harnessing the rich potential of open-end questionnaire responses.

6.4. Basic Research in Expert Opinion

There are two areas for basic research in expert opinion apparent from this study. One is the virtual absence of a critical scientific literature on expert opinion as mentioned in the introduction. The second is protocol analysis of human information processing in experts and nonexperts in soliciting expert opinion as developed from the results of this study. Each is briefly outlined.

The various scientific areas contributing to and concerned with expert opinion, as indicated in the introduction, have not effectively communicated with each other or benefitted from a cross-fertilization of method and findings. These partially overlapping areas include opinion polling, psychometrics, decision theory, problem solving, policy sciences, social and technological forecasting, and the various applications of expert opinion in law, clinical practice, management, defense, and other areas.

This recommendation is aimed at integrating method and findings in the scattered literature on expert opinion in these diverse areas for the first time, at setting up more rigorous scientific standards to minimize widespread abuses of expert opinion, and at promoting needed scientific advances to maximize the contribution of expert opinion to society.

The integration of the data on expert opinion should meet four goals. First, evaluate the methodology and findings of current use

of expert opinion in government, industry and science. Second, develop a conceptual framework for the field as a whole in the form of a rudimentary theory of the social use of expert opinion. Third, generate recommendations for future research in expert opinion, with special emphasis on methodological improvements for greater amenability to scientific method. And finally, provide recommendations for more cost-effective use of expert opinion and for key safeguards against major abuses in practice. It is hoped that such a study would contribute toward the emergence of an applied scientific field in its own right, marked by a critical and self-corrective literature on the social use of expert opinion.

Turning to the second recommendation, protocol analysis simply refers to the notion of obtaining data at the time experimental behavior is observed. This is done in participatory polling when the subject is asked to write out his justification for responding to an item with his particular rating. The adversary reason is the item protocol. This procedure has the great advantage that the respondent tells us what he is thinking at the time he makes his judgment. This approach is obviously superior to analysis from hindsight without the benefit of the "on-the-spot" protocol. The general intent of this second recommendation is that basic research should be performed on the information processing that occurs when experts respond to opinion questionnaires, preferably by more sophisticated forms of protocol analysis than those introduced in participatory polling.

The problem of snappy answers to complex problems was raised in the interpretation section for lengthy questionnaires where many questions are answered in a relatively short time. As response time becomes shorter and shorter, the quality of opinion is likely to get poorer and poorer. Human information processing time requires certain minima--granting large individual differences. Simple and disjunctive reaction times require only fractions of a second; simple verbal responses, as in ordinary conversation, require seconds; complex chains of reasoning require minutes or hours.

Protocol analysis could involve clocking response times for items, or asking subjects to "think out loud" in arriving at their opinion, following a structured procedure. Samples of such behaviors could be obtained from individual experts as part of a study on participatory polling to reconstruct patterns of expert information processing in arriving at opinions. Such studies could lead to improved types of opinion protocols for more effective expert opinion and to deeper understanding of the differences between experts and nonexperts.

Protocol analysis is but one example of basic research with expert opinion. It is rather important and especially needed since the item protocol is the distinguishing feature of participatory polling compared to conventional questionnaires, and it has not been systematically researched. Advances in this area could have major impact on the use of open-end responses in questionnaire instruments broadly considered. The proposed literature review to integrate the field of expert opinion, and protocol analysis of expert information processing appear to be promising initial steps toward scientific study of the social use of expert opinion.

7. SUMMARY AND CONCLUSIONS

The need for this study arose from two sources. First, prior review of Air Force and industrial practices in long-range R&D planning indicated that the most desirable and useful types of planning requires substantial participation of planners and decisionmakers in soliciting their assumptions, values, and opinions. Second, current techniques for obtaining such inputs, such as Delphi, suffer from numerous and severe limitations that need to be overcome. In response to these needs, the technique of participatory polling was developed for a prototype trial in this study. This polling technique has five key features: 1) iterative polling, 2) participant-interpreted reasons for responses, 3) evaluation of the interpretations of others, 4) quantitative and qualitative group feedback, and 5) formal questionnaire design and analysis.

7.1. Method

Participatory polling was applied to a sample of 10 Rand experts for long-range R&D planning for the close air support mission in the Air Force. A questionnaire of 24 items was developed, requesting importance ratings and reasons for ratings of various general scenarios concerned with: 1) type of war, 2) finding fixed versus mobile targets, 3) weather conditions, and 4) manned versus unmanned air penetration. Ratings and reasons were collected in the first round, and fed back anonymously to the group for the final second round results. The basic experimental design was embodied in a three-way analysis of variance consisting of three main effects and their associated interactions (i.e., 10 subjects versus 24 items versus two rounds).

7.2. Findings

The data were computerized and subjected to extensive quantitative and qualitative analyses, with the following seven major findings.

1. Differences Between Mission Requirements: Item differences, which represent differences in mission requirements, were substantial, credible, and statistically reliable (e.g.,

mobile targets were generally considered more important than fixed targets, and missiles were viewed as a greater threat than guns for close air support planning).

2. Test Reliability: Test-retest reliability under round one versus round two conditions (approximately one month apart) was quite high, with a median item reliability coefficient of .89 for the 24 items (product-moment correlations). Only about one-sixth of all item ratings (240 for all subjects combined) were changed from round one to round two. A mild convergence effect was noted with almost all item standard deviations decreasing from round one to round two, with a median reduction of 15 percent.
3. Disagreement Between Experts: Expert differences were substantial in evaluating the overall importance of items, and were statistically reliable. For example, there was considerable disagreement among experts over the value of manned versus unmanned aircraft for close air support planning, and over the impact of weather and visibility conditions.
4. Agreements Between Experts: Concurrence of expert opinion was strikingly demonstrated when specific reasons for opinions were rated. That is, experts consistently agreed among themselves in the second round as to whether particular reasons collected from the first round were important or unimportant within the framework of a given item.
5. Personalized Expert Opinion: A related finding is that the Rand experts held unique, specialized, and issue-specific opinions demonstrated by statistically reliable results for the interaction between items and subjects, and by personalized opinions that generally remained stable from the first to the second round. The observed specialization of opinion was further supported by statistically reliable discrimination of the importance of reasons offered in support of opinions.
6. Appreciation of Expert Opinion: In the qualitative findings, the most frequently reported insight gained from this polling

experience was awareness of the rather wide range of observed opinion and the highly diversified set of reasons and assumptions offered to support such opinion in the relatively specialized area of close air support.

7. Usefulness of Formal Technique: Almost all participants found the experimental procedure useful for long-range R&D planning, and provided numerous recommendations for improving quantitative and qualitative procedures.

7.3. Conclusions

These initial prototype results are most encouraging. Results from expert opinion on R&D planning for the close air support mission were generally credible and internally consistent. Theoretical leads into similarities and differences of expert opinion, and in the individualized structure and stability of expert opinion were notable. Formal experimental procedures are explicit and replicable; they represent a balanced form of adversary procedure, amenable to quantitative treatment; and associated measures were consistently found to be statistically reliable. The overall method appears to be more cost-effective and more scientifically defensible than competing iterative polling techniques.

REFERENCES

AMERICAN PSYCHOLOGICAL ASSOCIATION
Standards for Educational and Psychological Tests and Manuals
Washington, D. C., 1974.

ANASTASI, A.
Psychological Testing
Third Edition, Macmillan, New York, 1968.

BEDFORD, MICHAEL T.
The Future of Communications Services into the Home
Bell Canada Business Planning, September 1972, (proprietary).

BERELSON, B., and G. A. STEINER
Human Behavior: An Inventory or Scientific Findings
Harcourt, Brace, and World, New York, 1964.

BORKO, H.
A Study of the Needs for Research in Library and Information Science Education
Institute of Library Research, University of California, Los Angeles, California, 1970.

BRODEUR, PAUL
"Annals of Industry--Casualties of the Workplace"
New Yorker, November 12, 1973.

CAMPBELL, R. M.
A Methodological Study of the Utilization of Experts in Business Forecasting
Ph.D. Thesis, University of California, Los Angeles, 1966.

CANTRIL, HADLEY
"The Prediction of Social Events"
Journal of Abnormal and Social Psychology, Vol. 33, 1938, pp. 364-389.

DERIAN, JEAN-CLAUDE and FRANCIOSE MORIZE
Delphi in the Assessment of Research and Development Projects
Futures, October 1973, pp. 469-483.

FELDMAN, PHILIP
"Internal and External Delphi Panel Comparison"
Delphi: The Bell Canada Experience, Bell Canada, 1972.

GOLDBERG, L. R.
"Simple Models or Simple Processes? Some Research on Clinical Judgments"
American Psychologist 23 (1968), pp. 483-496.

GORDON, T. J., and O. HELMER
Report on a Long-Range Forecasting Study
The Rand Corporation, P-2982, September 1964.

HAYDEN, SPENCER
"How Industry is Using Technological Forecasting"
Management Review, May 1970, pp. 4-15.

HELLER, FRANK A.
"Group Feedback Analysis: A Method of Field Research"
Psychological Bulletin, Vol. 72, No. 2, pp. 108-117, 1969.

HERSCH, CHARLES
"The Discontent Explosion in Mental Health"
American Psychologist, 1969, pp. 597-606.

KAPLAN, A., A. L. SKOGSTAD, and M. A. GIRSCHICK
"The Prediction of Social and Technological Events"
Public Opinion Quarterly, Vol. 14, No. 1, 1950, pp. 13-110.
(See also The Rand Corporation, P-93, 1949.)

MCGREGOR, DOUGLAS
"The Major Determinants of the Prediction of Social Events"
Journal of Abnormal and Social Psychology, Vol. 33, 1938, pp. 179-204.

OSKAMP, S.
"Overconfidence in Case-Study Judgments"
Journal of Consulting Psychology 29, 1965, pp. 261-265.

PRESS, S. JAMES and ALVIN J. HARMAN
Methodology for Subjective Assessment of Technological Advancement
The Rand Corporation, R-1375, 1975.

REISMAN, A., S. J. MANTEL, JR., B. V. DEAN, and N. EISENBERG
Evaluation and Budgeting Model for a System of Social Agencies
Department of Operations Research, Case Western Reserve University, Technical Memorandum No. 167, 1969.

SACKMAN, HAROLD
Delphi Critique
D. C. Heath, Lexington, Mass., 1975.

SHERIF, M.
The Psychology of Social Norms
Harper, New York, 1936.

SILLS, DAVID L. (Ed.)
International Encyclopedia of the Social Sciences
Macmillan, New York, 1968.

SLOVIC, PAUL
"Psychological Study of Human Judgment"
The Journal of Finance, September 1972, pp. 779-798.

THORNDIKE, ROBERT L.
Personnel Selection
Wiley, New York, 1949.

WIRT, JOHN G., A. J. LEIBERMAN, and ROGER LEVIEN
R&D Management: Methods Used by Federal Agencies
The Rand Corporation, R-1156-HEW, January 1974.

ZARNOWITZ, VICTOR
"On the Accuracy and Properties of Short-Term Economic Forecasts"
The Task of Economies, Forty-Fifth Annual Report of the National
Bureau of Economic Research, 1965.

Exhibit A

FIRST ROUND QUESTIONNAIRE INSTRUCTIONS

1. INTRODUCTION

These forms are part of a Project RAND study on long-range R&D planning for USAF. This pilot study is concerned with comparative evaluation of planning requirements and candidate projects for the Close Air Support mission. The intent of the overall task ~~is~~ to systematically assess 1) the relative priorities of competing mission requirements, 2) preferences for proposed projects to satisfy these requirements, and 3) the rationale for priorities and preferences as determined by opinions expressed by Air Force policy makers and planners.

This questionnaire is focused on the first of those three objectives--the relative priorities of competing mission requirements. Since the determination of such relative priorities is a matter of judgment (as opposed to being amenable to quantitative analysis), we are soliciting your systematic judgment along with that of other qualified experts. The results of this exercise will furnish a composite set of priority-ordered operational needs, and the rationale for such needs, in the close air support mission. These results will serve as one of several criteria in the selection among alternative candidate R&D projects.

2. GENERAL INSTRUCTIONS

Your individual responses will be kept strictly anonymous during the course of this study and in the final report. The iterative polling procedure will permit you to become fully familiar with the anonymous opinions and reasons for opinions given by the other participants in the study. Although this procedure may resemble Delphi in certain respects, it represents a substantial modification in questionnaire design, development, and application.

The general procedure involves two rounds of the questionnaire. In the first round, your opinion is solicited on the relative importance of several operational problem areas in the Close Air Support mission. The scale runs from 0 = unimportant, to 1 = marginally important, to 2 = important, with 3 = very important as the highest rating.

The next step is crucial for this technique. After deciding upon your rating, you are requested to record the main reason why you chose the particular importance rating for each item. The "main reason" should be as terse as possible, hopefully a phrase or a single sentence. The distinguishing feature of this questionnaire procedure is that opinion is backed up by reasons. Instead of the investigators attempting to interpret the opinion of others, this procedure is designed to have participants interpret their own opinions. Previous studies have shown that experts and policy makers are particularly interested in the variety and credibility of reasons for and against expressed opinion so that more enlightened choices may be made. In order for this goal to be accomplished, it is essential that every importance rating be accompanied by a reason for the rating.

In the second round, the results of the first round are made available to all participants. That is, you are informed of the full frequency distribution of ratings by all participants, with a listing of reasons for the importance ratings for each item. After inspecting group feedback and becoming familiar with the arguments and rationale for the observed spread of group opinion, you are again asked to provide an importance rating for the item, using the same scale.

Each participant will have the benefit of seeing whether the group leans toward consensus or polarization, and the reasons offered by the group for pooled results. In addition, each participant will be able to determine the extent to which his anonymous opinions match those of others in the questionnaire sample.

It should be noted that this procedure is deliberately designed to elicit differences in opinion, and the reasons for such differences to facilitate the generation of a range of alternatives from which the policy maker and planner can choose. It is not designed as a tool for generating consensus through group suggestion, nor does it put a premium on consensus as is the case with Delphi. The approach is fundamentally that of a balanced adversary process as determined by participant feedback.

The value of this study depends in large part upon the quality of your responses. If any ambiguities arise, check with the research representative who arranged your participation in this project.

3. GENERAL APPROACH

There are three broad categories of operational problems in the close air support mission that should be amenable to some improvement through technological advances:

(1) Coping with defenses: evading them, degrading them through use of countermeasures, suppressing them through direct attack, or physically destroying them.

(2) Finding and identifying targets: the entire process of transferring target information to the person (pilot) who is controlling the ordnance delivery.

(3) Killing the target: delivering adequate amounts of appropriately-designed ordnance with the necessary accuracy.

In an attempt to simplify the procedural problems as much as possible in this initial experiment, we have made the assumption that today our most serious problems are in the areas of finding targets and coping with defenses. Given some success at those tasks, we now have weapons and delivery systems that are quite effective against a large class of close support targets.(1) We therefore concentrate primarily on the first two classes of operational problems listed above. In this questionnaire we are seeking some understanding of the relative priorities between those problem areas, which will then be used in the process of evaluating the relative merit of competing R&D proposals.

---

(1) One can argue, of course, that the three classes of problems are so interrelated that progress in any one area contributes to success in the other areas. While that is certainly true to a degree, we believe that they can be separated sufficiently to make this kind of analysis useful.

A level of detail is sought that will give useful and acceptable guidance to the R&D decisionmaker. In this first experiment we have elected to subdivide each broad problem area into two parts:

(1) coping with defenses;

--- coping with guided defenses (i.e., missiles with post-launch guidance)

--- coping with unguided, proliferation weapons (i.e., guns)

(2) finding and identifying targets;

--- fixed targets

--- movable or mobile targets.

We justify this subdivision on the basis that the technological advances that might be devised to help solve those problems are, in broad terms, different between the subdivisions in each problem area, and therefore, this subdivision should be useful to the R&D planners.

We have further defined three "operational context" parameters:

(1) scenario (two types)

(2) visibility (two levels)

(3) degree of penetration (two classes)

These are defined in some detail below. We therefore have a potential matrix of 2^5 entries; (2 kinds of defenses) x (2 kinds of targets) x (2 scenarios) x (2 levels of visibility) x (2 degrees of penetration). However, by deleting some contradictory combinations, the final matrix contains 24 rather than 32 entries.*

Definition of Terms

Scenario: No one, of course, knows exactly where, or under what rules of engagement, the next wars will be fought. In fact, underlying each

*The items deleted are those concerned with coping with defenses in situations involving penetration of defenses with unmanned vehicles. Although it may be argued that survival of unmanned systems over the battlefield is not a trivial problem, it was deleted from this initial, experimental questionnaire in the interest of simplicity.

decision in R&D project selection, there is a (usually implicit) set of opinions on the relative likelihood and importance of the various alternatives. At this stage in the project, we wish to make this part of the decision process more visible by specifying two broad classes of combat environment and asking for your judgments on the importance of certain operational problems separately in each environment.

The two environments selected here are NATO and Third Area, defined as follows:

a. NATO: A major nonnuclear engagement in central Europe sometime during the mid-1980s, probably along the lines of the classic scenario; a massive Pact armored thrust focused on a quick conquest of territory, possibly all the way to the Rhine. Full utilization of each side's forces and equipments short of nuclear weapons.

b. Third Area: An engagement of major importance, involving forces of the U.S. and the S.U., but in a "third area" so that neither power is directly defending its own borders. The Mid-East is currently a likely spot for such a confrontation, but by the 1980s, the focus could shift to the Indian subcontinent, to South America, etc. The exact location is not too important. The engagement would likely involve a much more cautious and controlled application of force by each major power than in the NATO scenario, but first-line systems would be employed. The war could be locally intense, and could last for months or even years, but the average rate of total resource allocation would probably be less than in the NATO case.

Visibility: The problems of finding and identifying targets, and of coping with defenses, are both influenced to an important degree by the visibility conditions existing at the time. The options open to both sides under clear day conditions are very different from those available as the visibility begins to degrade, either through atmospheric moisture, smoke/haze, or darkness.

We have elected the simplest possible categorization for this initial experiment; clear day and restricted visibility.

a. Clear day; essentially unrestricted visibility and ceiling, during daylight hours.

b. Restricted visibility; an intermediate condition, such as might be expected during winter in central Europe when weather hampers flight operations but does not cause complete stopping of all ground activity. Typically, visibility might be less than a mile and ceilings less than 2000 ft, but these numbers should not be treated as anything more than very general indications.

Degree of Penetration: Back in the old days, when there was only free-fall ordnance, close air support operations required at least some exposure to enemy surface-based defenses. Today, there is at least the option to consider a set of possibilities in which manned air vehicles are never exposed to battlefield defenses, but instead stand back and launch missiles, RPVs, etc. against the targets. Therefore, two cases are considered; penetration and nonpenetration.

a. Penetration; a manned aircraft flies directly over the battlefield or at least near enough to come within effective coverage of surface-based defenses that the enemy might deploy in the forward battle area.

b. Nonpenetration; all manned vehicles stay sufficiently well back over friendly territory so that they are not subject to effective fire from enemy surface-based defenses.

Note that these assumptions completely ignore the problem of coping with enemy fighter aircraft. This is comprehensively treated under separate analysis and is not considered here.

4. INSTRUCTIONS FOR PLANNING REQUIREMENTS QUESTIONNAIRE

This questionnaire lists 24 combinations of operational problems and environmental context.

Environmantal Context	Operational Problems
Type of Environment	Finding Target
1. NATO	7. Fixed Target
2. Third Area	8. Mobile Target
Visibility Conditions	Coping with Defenses
3. Clear Day	9. Guided Weapons
4. Restricted Visibility	10. Guns
Penetration	
5. Manned Penetration	
6. Nonpenetration	

You are asked to rate each of the 24 items with respect to the importance scale as described in the previous section. In addition, you are requested to write in the main reason for your importance rating in the space provided for each item. As mentioned earlier, your written reason should not be longer than a single sentence. Please use pencil; it will be easier for you to change your own responses as required.

A final word on what is meant by "importance." How critically we need some improvement in a particular operational area is a combination of two factors; the "absolute" importance of that operational capability, and the extent to which we can accomplish it with today's forces. Consider the following matrix:

		Today's Capability	
		Hi	Low
Importance of Mission	Hi	B	A
	Low	C	B

Items falling in box "A" would deserve the highest importance rating, those in box "C" the lowest rating, and those in boxes "B" an intermediate rating. We are not asking you to decompose your rating in this manner (although some such notions may emerge in your 'reason' behind your rating); this example is used only as a way of further clarifying the importance rating.

Exhibit B

SECOND-ROUND QUESTIONNAIRE WITH DESCRIPTIVE STATISTICS

1. INSTRUCTIONS

This questionnaire contains two forms for the second round. We would greatly appreciate your filling out these forms promptly and returning them to us within one week of your receipt of this material. The second round forms have been designed to be self-administering. However, if any problems arise, please do not hesitate to call us.

The first form is the Second Round Evaluation of Close Air Support Mission Requirements. This form incorporates the 24 items of the first-round questionnaire with new information. The new material includes statistical and verbal group feedback from the first round. The descriptive statistics include the rating scale frequency distribution, the mean, and standard deviation for each item.

You are asked to examine the listed reasons carefully for each item, and then to provide the evaluative information requested for the item. All first-round ratings and reasons have been kept anonymous. Rate the importance of each of the reasons put forth for assessing each item, until all listed reasons are evaluated with the standardized rating scale. (Very Important = 3; Important = 2; Marginally Important = 1; Unimportant = 0.) After you have examined and rated the several reasons listed for the item, reconsider your final overall importance rating for the R&D value of the item for the operational Close Air Support Mission.

Your second rating may either remain the same or be revised in accordance with your estimate of group feedback in relation to your personal opinion. As mentioned in the instructions for the first round, this procedure is deliberately designed to elicit differences in opinion, and the reasons for such differences, to generate a range of competing alternatives from which the policy maker and planner can choose. It is deliberately based on a balanced adversary procedure, and is not intended to induce consensus through group suggestion.

You are requested to rate group feedback on all 24 items, and to provide overall ratings for each item as indicated above for R&D planning for the Close Air Support Mission. Space is also provided for

additional, optional comments on each item in the event you feel some important or significant consideration has been omitted or overlooked.

The second form is a quick, one-page Questionnaire Evaluation. The items are concerned with your overall estimate of the benefits, liabilities, and general impact of this technique in support of R&D planning. The questions in this form are self-explanatory. Open-end items should not be longer than a single sentence.

A copy of your original first-round questionnaire is also enclosed for your convenience in looking up definitions, procedures, and your first-round responses if you so desire.

Please fill in your name on the third page, and use pencil to make it easier for you to change any entries. Your second-round responses will also be kept anonymous in the final results for this exercise.

We greatly appreciate your helpful and constructive cooperation in developing and evaluating this technique for more effective R&D Air Force planning.

Name ______________________

2. SECOND ROUND EVALUATION FOR CLOSE AIR SUPPORT MISSION

1. Coping with Missiles: NATO War, Clear Day, Manned Penetration

First round frequencies of importance ratings:

0 - Unimportant 0   2 - Important 4   Mean 2.40
1 - Marginally Important 1   3 - Very Important 5   St.D. .70

	Reasons for Ratings	Importance Rating for Each Reason	
a.	SAM threat and general missile threat is a vital problem for NATO manned penetration.	2.7	.48
b.	This operational requirement is very high priority for operators of manned systems.	2.6	.70
c.	ECM and missile tactics are reasonably effective today.	1.2	.92
d.	Only shallow and reduced close air support will be tolerable against heavy expected losses.	1.6	1.43
	*Your Second Round Importance Rating for this item.	2.6	.52
		(Means)	(SD's)

Additional Second Round Reason: ______________________

____

2. Coping with Missiles: Third Area, Clear Day, Manned Penetration

First round frequencies of importance ratings:

0 - Unimportant 0   2 - Important 4   Mean 1.80
1 - Marginally Important 4   3 - Very Important 2   St.D. .79

	Reasons for Ratings	Importance Rating for Each Reason	
a.	ECM and missile tactics are reasonably effective.	1.5	1.18
b.	It is especially important to cope with high altitude missiles to avoid low-altitude proliferation weapons.	2.2	.79
c.	Political pressures require minimal manned penetration losses.	1.8	1.03
d.	Depends on area and availability of missiles--more important for Middle-East than other areas.	1.7	1.16
	*Your Second Round Importance Rating for this item.	2.0	.67
		(Means)	(SD's)

Additional Second Round Reason: ______________________

____

*Overall mean and standard deviation for the item importance rating is listed at this entry throughout this questionnaire.

3. Coping with Guns: NATO War, Clear Day, Manned Penetration

First round frequencies of importance ratings:

0 - Unimportant 2   2 - Important 5   Mean 1.50
1 - Marginally Important 2   3 - Very Important 1   St.D. .97

Reasons for Ratings	Importance Rating for Each Reason	
a. Not important because a solution exists: overflight.	1.1	1.20
b. High priority because large numbers of guns are concentrated along FEBA, with current lack of adequate countermeasures.	1.9	1.29
c. USAF already has capability to counter guns, especially if standoff capability is available.	1.1	1.10
Your Second Round Importance Rating for this item.	1.5	.97
	(Means)	(SD's)

Additional Second Round Reason: ____________________

4. Coping with Guns: Third Area, Clear Day, Manned Penetration

First round frequencies of importance ratings:

0 - Unimportant 3   2 - Important 3   Mean 1.20
1 - Marginally Important 3   3 - Very Important 1   St.D. 1.03

Reasons for Ratings	Importance Rating for Each Reason	
a. Not important because big caliber radar-controlled guns tend to be the exception: overflight is the solution.	1.2	1.14
b. Gun concentration is high and there is a current lack of adequate countermeasures.	1.5	1.18
c. Experience in Vietnam and Yom Kippur Wars indicate that better manned system capability is required for this threat.	1.6	1.35
d. Not important if gun concentration is not dense and if standoff capability is available.	1.8	1.40
Your Second Round Importance Rating for this item.	1.3	1.06
	(Means)	(SD's)

Additional Second Round Reason: ____________________

5. Finding Fixed Targets: NATO War, Clear Day, Manned Penetration

First round frequencies of importance ratings:

0 - Unimportant 0	2 - Important 6	Mean 1.60
1 - Marginally Important 4	3 - Very Important 0	St.D. .52

Reasons for Ratings	Importance Rating for Each Reason	
a. We now have good capability for finding and identifying fixed targets in clear weather.	1.7	1.34
b. Because of effective enemy defenses, manned systems will be forced to use tactics not optimum for target detection.	1.9	1.10
c. There are significant problems with common grid systems and precision navigation systems.	2.1	.88
Your Second Round Importance Rating for this item.	1.5	.53
	(Means)	(SD's)

Additional Second Round Reason: ____________________

6. Finding Fixed Targets: Third Area, Clear Day, Manned Penetration

First round frequencies of importance ratings:

0 - Unimportant 0	2 - Important 6	Mean 1.80
1 - Marginally Important 3	3 - Very Imporant 1	St.D. .63

Reasons for Ratings	Importance Rating for Each Reason	
a. Because of effective enemy defenses, manned systems will be forced to use tactics not optimum for target detection.	1.6	1.08
b. There are very few fixed targets for which good capability does not exist in clear weather, particularly if locations are known.	1.7	1.16
c. There are significant problems with common grid systems and precision navigation systems.	2.1	.74
d. Important problem because locations are less well known than in NATO, and natural cover may present greater difficulties.	1.7	.82
Your Second Round Importance Rating for this item.	1.8	.63
	(Means)	(SD's)

Additional Second Round Reason: ____________________

7. Finding Mobile Targets: NATO War, Clear Day, Manned Penetration

First round frequencies of importance ratings:

0 - Unimportant 0	2 - Important 3	Mean 2.30
1 - Marginally Important 2	3 - Very Important 5	St.D. .82

	Reasons for Ratings	Importance Rating for Each Reason	
a.	We must stop tank assault or lose; we have trouble now and camoflage makes it tougher.	2.6	.97
b.	Under clear conditions in a NATO setting there should be no problem finding mobile targets.	1.2	1.14
c.	Effective defenses and navigational difficulties create many tactical problems for CAS.	2.2	1.03
d.	Demands of time-urgent allocation of scarce sorties require improved real-time battlefield surveillance.	2.4	.97
	Your Second Round Importance Rating for this item.	2.5	.53
		(Means)	(SD's)

Additional Second Round Reason: ______________________________

____

8. Finding Mobile Targets: Third Area, Clear Day, Manned Penetration

First round frequencies of importance ratings:

0 - Unimportant 0	2 - Important 2	Mean 2.60
1 - Marginally Important 1	3 - Very Important 7	St.D. .70

	Reasons for Ratings	Importance Rating for Each Reason	
a.	With sorties at a premium, we can't afford to waste any.	2.0	.82
b.	No problem finding mobile targets under clear day conditions in third area.	.70	1.06
c.	Great difficulties in identifying enemy targets when both sides may be using same type of tank.	2.3	.82
d.	Mobile targets may be hard to detect, depending on terrain and camoflage.	2.4	.84
	Your Second Round Importance Rating for this item.	2.7	.48
		(Means)	(SD's)

Additional Second Round Reason: ______________________________

____

9. Coping with Missiles: NATO War, Restricted Visibility, Manned Penetration

First round frequencies of importance ratings:

0 - Unimportant 0	2 - Important 3	Mean 2.30
1 - Marginally Important 2	3 - Very Important 5	St.D. .82

Reasons for Ratings	Importance Rating for Each Reason	
a. The urgent need for CAS means that a/c must operate even under low ceilings and at short slant ranges if standoff capability is not available.	2.2	1.23
b. High priority because of quantity and quality of SAM's and AIM's, and associated radar which overcomes effect of restricted visibility and forces a/c to lower altitudes.	2.1	.88
c. Important mission, however ECM and tactics are reasonably effective.	1.1	1.10
d. Although coping with high-altitude missiles is important, technological requirements are not that difficult.	1.0	.67
Your Second Round Importance Rating for this item.	2.6	.52
	(Means)	(SD's)

Additional Second Round Reason: ______________________________

____

10. Coping with Missiles: Third Area, Restricted Visibility, Manned Penetration

First round frequencies of importance ratings:

0 - Unimportant 1	2 - Important 3	Mean 1.60
1 - Marginally Important 4	3 - Very Important 2	St.D. .97

Reasons for Ratings	Importance Rating for Each Reason	
a. Forget the sortie if missiles are that much of a problem.	1.4	1.43
b. High priority because of effective missile systems and less effective counter-missile systems with restricted visibility.	1.8	1.14
c. Important mission, however ECM and tactics are reasonably effective.	1.5	1.08
d. With exception of Middle East, third areas are not likely in the near future to have sophisticated SAM systems.	.70	.68
Your Second Round Importance Rating for this item.	1.7	.95
	(Means)	(SD's)

Additional Second Round Reason: ______________________________

____

11. Coping with Guns: NATO War, Restricted Visibility, Manned Penetration

First round frequencies of importance ratings:

0 - Unimportant 0	2 - Important 7	Mean 1.70
1 - Marginally Important 3	3 - Very Important 0	St.D. .48

	Reasons for Ratings	Importance Rating for Each Reason	
a.	If mission must be flown, accept the risk with current capability.	.70	1.06
b.	USAF capabilities against gun defenses are probably not as good with restricted visibility.	1.5	.85
c.	Overflight is a current solution requiring no research.	1.0	1.16
d.	Must fly lower to see targets, driving a/c into gun and optical range.	2.6	.70
	Your Second Round Importance Rating for this item.	1.8	.63
		(Means)	(SD's)

Additional Second Round Reason: ______________________________

12. Coping with Guns: Third Area, Restricted Visibility, Manned Penetration

First round frequencies of importance ratings:

0 - Unimportant 1	2 - Important 6	Mean 1.50
1 - Marginally Important 3	3 - Very Important 0	St.D. .71

	Reasons for Ratings	Importance Rating for Each Reason	
a.	USAF capabilities against gun defenses need improvement for restricted visibility conditions which forces a/c into optical range.	2.1	.88
b.	This threat is less severe than in clear weather because of difficulty AA has in acquiring a/c target.	1.8	1.23
c.	Overflight is the solution--the research burden is on target finding and attacking.	1.1	1.10
d.	Importance of this threat depends on concentration of guns in third area and whether standoff capability is sensitive to restricted visibility.	1.8	.92
	Your Second Round Importance Rating for this item.	1.6	.70
		(Means)	(SD's)

Additional Second Round Reason: ______________________________

13. Finding Fixed Targets: NATO War, Restricted Visibility, Manned Penetration

First round frequencies of importance ratings:

0 - Unimportant 0	2 - Important 4	Mean 1.80
1 - Marginally Important 4	3 - Very Important 2	St.D. .79

Reasons for Ratings	Importance Rating for Each Reason	
a. Very important research area because of the desire to overfly the weather to avoid defenses.	1.7	.95
b. Send a missile or RPV for this threat, not manned a/c.	1.8	1.40
c. Assuming air base attacks, use special systems for few high priority fixed targets that cannot wait.	2.0	1.05
d. With restricted visibility there is a lack of ability to accurately locate targets, except for those that radiate.	1.8	1.03
Your Second Round Importance Rating for this item.	1.6	.70
	(Means)	(SD's)

Additional Second Round Reason: ______________________________

____

14. Finding Fixed Targets: Third Area, Restricted Visibility, Manned Penetration

First round frequencies of importance ratings:

0 - Unimportant 1	2 - Important 5	Mean 1.80
1 - Marginally Important 2	3 - Very Important 2	St.D. .92

Reasons for Ratings	Importance Rating for Each Reason	
a. Unimportant fixed targets can be attacked when convenient, i.e., clear weather.	1.5	1.35
b. Very important research area because of the desire to overfly the weather to avoid defenses.	1.4	.84
c. Send a missile or RPV for this threat, not manned a/c.	1.2	1.40
d. We lack ability to accurately locate targets, except for those that radiate.	1.8	.92
Your Second Round Importance Rating for this item.	1.6	.84
	(Means)	(SD's)

Additional Second Round Reason: ______________________________

____

15. Finding Mobile Targets: NATO War, Restricted Visibility, Manned Penetration

First round frequencies of importance ratings:

0 - Unimportant 0	2 - Important 3	Mean 2.50
1 - Marginally Important 1	3 - Very Important 6	St.D. .71

	Reasons for Ratings	Importance Rating for Each Reason	
a.	With restricted visibility, ground war tends to slow down and fewer sorties are used, therefore, CAS is not decisive for this threat.	.80	.63
b.	Demands of time-urgent allocation of scarce sorties require near-real-time battlefield surveillance, especially for NATO/CAS.	2.7	.48
c.	Extremely important to find mobile targets in marginal weather, which occurs much of the time in NATO.	2.7	.48
d.	We must stop tank assault or lose.	2.3	1.06
	Your Second Round Importance Rating for this item.	2.6	.52
		(Means)	(SD's)

Additional Second Round Reason: ______________________

____

16. Finding Mobile Targets: Third Area, Restricted Visibility, Manned Penetration

First round frequencies of importance ratings:

0 - Unimportant 1	2 - Important 4	Mean 2.10
1 - Marginally Important 1	3 - Very Important 4	St.D. .99

	Reasons for Ratings	Importance Rating for Each Reason	
a.	Less urgent than in NATO, however, mobile targets may be much less visible, depending on terrain.	2.3	.68
b.	With inaccurate location and lengthy processing time, numbers and value of effective targets are decreased.	2.0	.82
c.	We must stop tank assault or lose.	1.3	1.16
d.	Important research area because of desire to overfly the weather to avoid defenses.	1.6	.97
	Your Second Round Importance Rating for this item.	2.1	.88
		(Means)	(SD's)

Additional Second Round Reason: ______________________

____

17. Finding Fixed Targets: NATO War, Clear Day, Nonpenetration

First round frequencies of importance ratings:

0 - Unimportant 0	2 - Important 5	Mean 1.90
1 - Marginally Important 3	3 - Very Important 2	St.D. .74

	Reasons for Ratings	Importance Rating for Each Reason	
a.	You can't hit what you can't see; intelligence problem with fixed target identity and location.	1.7	1.12
b.	Unmanned navigational and terminal sensor capability required to replace manned penetration systems which are seriously degraded by effective defenses.	2.3	1.06
c.	Except for easily found fixed targets, research is not vital for developing standoff against other types of targets.	1.1	.99
d.	Present capability is good; this is a very small target class.	1.2	1.14
	Your Second Round Importance Rating for this item.	1.9	.74
		(Means)	(SD's)

Additional Second Round Reason: ______________________________

18. Finding Fixed Targets: Third Area, Clear Day, Nonpenetration

First round frequencies of importance ratings:

0 - Unimportant 0	2 - Important 4	Mean 1.80
1 - Marginally Important 4	3 - Very Important 2	St.D. .79

	Reasons for Ratings	Importance Rating for Each Reason	
a.	If job can't be done with manned a/c, it is probably not worth the expense to do it at all, especially with lower target densities.	1.0	.82
b.	Develop good missile or RPV navigation and terminal sensors for this job and leave manned a/c at home.	1.9	1.20
c.	Except for easily found fixed targets, research is not vital for developing standoff against other types of targets.	1.2	1.14
d.	Locations are not as well known as in NATO setting, and ground-to-air communications are not up to NATO standards, which makes this requirement more important.	2.1	.88
	Your Second Round Importance Rating for this item.	1.9	.57
		(Means)	(SD's)

Additional Second Round Reason: ______________________________

19. Finding Mobile Targets: NATO War, Clear Day, Nonpenetration

First round frequencies of importance ratings:

0 - Unimportant 0	2 - Important 4	Mean 2.00
1 - Marginally Important 3	3 - Very Important 3	St.D. .82

Reasons for Ratings	Importance Rating for Each Reason	
a. Important mission, but present capability is reasonably good and no vital research is needed.	.90	.74
b. Capability required to replace manned penetration systems which are seriously degraded by tough defenses getting tougher.	2.3	1.06
c. Marginal, unless there is a two-way data link for constant update on position so that target will be there when weapon arrives.	1.8	1.23
Your Second Round Importance Rating for this item.	2.3	.82
	(Means)	(SD's)

Additional Second Round Reason: ______________________

____

20. Finding Mobile Targets: Third Area, Clear Day, Nonpenetration

First round frequencies of importance ratings:

0 - Unimportant 0	2 - Important 3	Mean 1.70
1 - Marginally Important 5	3 - Very Important 2	St.D. .82

Reasons for Ratings	Importance Rating for Each Reason	
a. Resupply is a big problem in third area--strike targets with minimum waste of weapons.	1.6	.84
b. Important mission because of generally poor ground-to-air liason in third areas.	2.1	.88
c. Unless it is a Middle East scenario, there may not be enough targets, which places more stringent requirements on midcourse and terminal guidance.	1.5	.85
d. Given lower target densities in third areas, it is probably not worth the expense to do it at all if we can't do it with manned a/c.	1.4	1.27
e. Although research is important for identification, this mission is low priority.	1.2	.92
Your Second Round Importance Rating for this item.	1.6	.70
	(Means)	(SD's)

Additional Second Round Reason: ______________________

____

21. Finding Fixed Targets: NATO War, Restricted Visibility, Nonpenetration

First round frequencies of importance ratings:

0 - Unimportant 0	2 - Important 2	Mean 2.00
1 - Marginally Important 4	3 - Very Important 4	St.D. .94

Reasons for Ratings	Importance Rating for Each Reason	
a. If we can do this in clear weather first, it just may be enough.	1.5	1.08
b. Although research is needed to achieve full capability, this mission is low priority.	1.2	1.03
c. We lack the ability to accurately locate targets, except, possibly, for those that radiate--we can't hit what we can't see.	2.2	1.14
d. Develop good missile or RPV navigation and terminal sensors for this job, and leave manned a/c at home.	2.2	1.03
Your Second Round Importance Rating for this item.	2.1	.88
	(Means)	(SD's)

Additional Second Round Reason: ______________________________

22. Finding Fixed Targets: Third Area, Restricted Visibility, Nonpenetration

First round frequencies of importance ratings:

0 - Unimportant 1	2 - Important 5	Mean 1.80
1 - Marginally Important 2	3 - Very Important 2	St.D. .92

Reasons for Ratings	Importance Rating for Each Reason	
a. Important mission becuase of generally poor ground-to-air liason in third areas.	1.8	1.14
b. Wait for better weather--unimportant mission.	1.4	1.27
c. More difficult to accurately locate targets; although numbers and value of targets are decreased, this is still an important mission requiring research.	2.0	.67
d. Although research is needed to achieve full capability, this mission is low priority.	1.4	1.27
Your Second Round Importance Rating for this item.	1.7	.82
	(Means)	(SD's)

Additional Second Round Reason: ______________________________

23\. Finding Mobile Targets: NATO War, Restricted Visibility, Nonpenetration

First round frequencies of importance ratings:

0 - Unimportant 0 | 2 - Important 2 | Mean 2.00
1 - Marginally Important 4 | 3 - Very Important 4 | St.D. .94

	Reasons for Ratings	Importance Rating for Each Reason	
a.	Although present capability is moderately good, the need for improved real-time battlefield surveillance and the great importance of this mission makes it high priority for research.	2.2	1.23
b.	Although research is needed to achieve full capability, this mission is low priority except for limited areas.	1.1	.88
c.	Do the job in clear weather first before getting fancy.	1.5	1.18
d.	Capability is required to replace manned penetration systems which are seriously degraded by effective defenses.	2.1	.99
	Your Second Round Importance Rating for this item.	2.2	.79
		(Means)	(SD's)

Additional Second Round Reason: ____________________

____

24\. Finding Mobile Targets: Third Area, Restricted Visibility, Nonpenetration

First round frequencies of importance ratings:

0 - Unimportant 0 | 2 - Important 3 | Mean 1.70
1 - Marginally Important 5 | 3 - Very Important 2 | St.D. .82

	Reasons for Ratings	Importance Rating for Each Reason	
a.	In third areas there is more freedom in waiting for the weather to clear.	1.6	1.27
b.	Except for a Middle-East setting, there may not be enough targets, which places more stringent requirements on midcourse and terminal guidance.	1.3	1.06
c.	Although research is needed to achieve full capability, this mission is low priority except for limited areas.	1.1	.88
d.	With inaccurate location and long processing times, numbers and value of targets are decreased in this setting.	1.4	1.08
	Your Second Round Importance Rating for this item.	1.6	.70
		(Means)	(SD's)

Additional Second Round Reason: ____________________

____

3. OVERALL QUESTIONNAIRE EVALUATION

*1. What was your most important insight in working with this questionnaire? Great range and diversity of expert opinions and assumptions for R&D planning.

*2. What information or circumstances made this insight possible? Variety and scope of expert-elicited reasons shown for each of the questionnaire items.

3. Did this insight primarily occur: alone at the desk 7, or in group discussion 0, or both 2?

*4. What was your greatest difficulty in responding to this questionnaire? Various types of ambiguity in interpreting items.

*5. What is your recommendation for reducing or eliminating this difficulty? Reduction of item ambiguity by various means.

6. How reliable are your overall reported estimates and responses for this questionnaire? Very Reliable 2, Fairly Reliable 6, Marginally Reliable 1, Unreliable 0

7. How successful was this questionnaire in helping you to organize your thinking and planning for the Close Air Support mission? Very Successful 0, Successful 2, Marginally Successful 7, Unsuccessful 1

*Open-end responses, discussed in the report. Most popular or most frequent responses to each of these is paraphrased and summarized above.